AD NUMBER

AD820392

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Critical Technology; AUG 1967. Other requests shall be referred to Rome Air Development Center, ATTN: EMLI, Griffiss AFB, NY 13440.

AUTHORITY

RADC ltr dtd 17 Sep 1971

OPTIMUM DEMODULATION OF PM AND FM SIGNALS.

CORNELL AERONAUTICAL LAB INC BUFFALO NY

AUG 1967

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded, by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

OPTIMUM DEMODULATION OF PM AND FM SIGNALS

John G. Lawton

Cornell Aeronautical Laboratory, Inc.

ASTIA, GAFB, N.Y., 22 Sep 67-309

FOREWORD

The work described in this report was performed by Cornell Aeronautical Laboratory, Inc., Buffalo, New York, for Rome Air Development Center, Griffiss Air Force Base, New York, during the period 14 April 1966 to 27 May 1967. The work was authorized under RADC Contract AF30(602)-4221 and was initiated under Project No. 4519, Task No. 451902.

RADC technical administration of this project was under the direction of the Signal Processing Section (EMCRS), Communications Research Branch, Communications Division. The project engineer was Mr. Alfred S. Kobos, Rome Air Development Center, EMCRS, Griffiss Air Force Base, N.Y. 13440.

At Cornell Aeronautical Laboratory, Dr. John G. Lawton was the project engineer. Major contributions were made by Drs. John T. Mleck, Sol Kaufman, John G. Lawton, and Mr. Harold D. Becker, under whose aegis the investigations reported in Chapters II, III, IV and V respectively, were performed.

Cornell Aeronautical Laboratory, Inc., has assigned Report No. UB-2254-B-1 to this document.

This technical report has been reviewed by the Foreign Disclosure Policy Office (EMLI). It is not releasable to the Clearinghouse for Federal Scientific and Technical Information in accordance with "Strategic Trade Control Program" and "International Traffic in Arms Regulation," AFR 400-10 and AFSCM 200-2.

This technical report has been reviewed and is approved.

Approved: Alfred S Kobos

ALFRED S. KOBOS
Project Engineer

Approved:

RICHARD M. COSEL
Colonel, USAF
Chief, Communications Division

FOR THE COMMANDER:

IRVING J. GABELMAN
Chief, Advanced Studies Group

ABSTRACT

This report presents the results of an investigation of the limiting performance attainable with angle modulation (PM and FM) systems. Estimators which implement the statistical criteria of minimum mean-square error (MMSE) and maximum a posteriori probability (MAP) estimation as well as advanced but physically realizable demodulators, such as phase-locked loops (PLL), are investigated and compared. Because of the nonlinear characteristic of angle modulation it is difficult to obtain analytic expressions for the performance of these estimators which are very useful at all values of input signal-to-noise ratio. However, separate asymptotic expansions are obtained for the large and the small signal-to-noise ratio regions. The large signal-to-noise ratio expansion is valid upward from a point below threshold and covers the range of signal-to-noise ratios of greatest practical interest. An asymptotic formula valid for all signal-to-noise ratios and small modulation indexes was derived; it is compatible with and overlaps the domains of the other expansions. The region which could not be effectively treated analytically was investigated by the use of computers. The threshold phenomenon in phase-locked loop receivers was investigated by means of a hybrid computer simulation. The results indicate that all, properly designed, demodulators for angle modulated systems attain the same improvement factor above threshold and that the major difference of practical importance is the location of the threshold. Factors which limit threshold extension in the reception of FM and the implementation of improved FM demodulators employing phase-locked loop structures are described.

SUMMARY

This report presents the results of investigations which were undertaken to establish the maximum performance attainable in the demodulation of random angle modulated (PM and FM) signals through the use of estimators based on certain statistical criteria of optimization. These criteria are the minimization of the mean-square error (MMSE), and the maximization of the a posteriori probability (MAP) (i.e. estimation of the modulation after observation of the noise perturbed signal).

Because angle modulation is an inherently nonlinear process it is difficult to obtain useful analytical expressions for the performance of these systems which are valid at all input signal-to-noise ratios and for all values of the modulation index. However, by a combination of analytic and computer techniques it has been possible to cover the required range of operating conditions. Asymptotic expansions for large and small input signal-to-noise ratios $(S/N)_i$ (ratio of the carrier power to noise power in the bandwidth of the modulation) have been obtained in Chapter II. One of these expansions applies to input signal-to-noise ratios such that, operation extends from a point below threshold on upward. The range of input signal-to-noise ratios and modulation indexes of most practical interest is covered by this result. It is believed that this constitutes the first quantitatively useful analytic derivation of the limiting performance for this region.

Another analysis presented in Chapter III examined the problem of minimum mean-square error estimation at very low input signal-to-noise ratios and obtained an asymptotic expansion valid in that domain for all values of the modulation index. For small values of modulation index an asymptotic expression valid at all input signal-to-noise ratios has been obtained in Chapter IV. This formula bridges the gap between the domains of the large and the small $(S/N)_i$ expansions at small values of the modulation index and is compatible with these asymptotic expansions.

There exists a region of intermediate $(S/N)_i$ and moderate to large modulation indexes which could not be treated adequately by analytic means. This region was investigated by the use of both digital and analog computers

(Chapters IV and V). The digital computer investigation yielded upper bounds (i.e. theoretical performance cannot be better than that computed) on the performance attainable for all values of $(S/N)_i$ and moderate values of the modulation index for PM modulation. The asymptotic behavior at high $(S/N)_i$ was found to be in agreement with the analytic expressions of Chapters II and IV. A large range, -10 to +30 db, of the intrinsic input signal-to-noise ratio $(S/N)_i$ was explored, but there was insufficient computer time available to explore a very large range of the modulation index. However, the detailed computer program is included (Appendix IV-A) and this should make it readily possible for the reader to extend the range of parameters investigated. Large modulation index FM systems, ratio of RF to modulation bandwidth up to 82 were investigated by the use of a hybrid computer (partly digital and partly analog). The results of this investigation, which was aimed primarily at determining the threshold performance of phase-locked loop (PLL) FM receivers are reported in Chapter V. The value of $(S/N)_i$ at which threshold occurs was determined as a function of phase-locked loop design and the degree of modulation employed. By use of the hybrid computer it was readily possible to carry out simulation of a zero delay phase-locked loop design suggested by the integral equation which describes maximum a posteriori estimation. The resulting threshold performance was determined rather accurately and it was found that by empirical modifications to the loop transfer function, the threshold characteristic could be improved further.

It is noted that previous PLL analyses based on a linearized model assume that the threshold occurs when the mean-square phase error predicted by that analysis reaches a nominal rms value of the order of 0.5 radians. Analysis and computer investigations show that it is possible to reduce the value of $(S/N)_i$ at which threshold occurs below that predicted by the above model. While this conclusion was reached without postulating any specific demodulator configurations it was also demonstrated empirically that the threshold of a phase-locked loop can be reduced below this value. Some recently published controversies regarding, for example, the effect of loop bandwidth and modulation on threshold performance, have been resolved.

Chapter I presents a discussion and comparison of the results of all of the investigations summarized above. Chapter II-V contain the analytical derivations and detailed descriptions of the computer simulations.

TABLE OF CONTENTS

Table of Contents (Con't.) Page No.

LIST OF FIGURES

LIST OF FIGURES

EVALUATION

This study was undertaken to determine the limitations of frequency-modulation receivers, insofar as the FM threshold is concerned. The contractor, Cornell Aeronautical Laboratory, was successful in doing this by employing a variety of methods, some analytical and some involving the use of simulation on a computer.

This report presents a consistent approach to the study of the threshold phenomenon in FM demodulation. For the first time, reasonable bounds have been set on the achievable threshold, determining the actual amount of improvement which can be expected for today's receivers.

The contractor has dealt with a very difficult and heretofore unsolved analytical problem. Because of the non-linearity of the equations involved, obtaining solutions was an extremely difficult task. The fact that the differing methods employed to achieve a solution lead to results which agree with each other, indicates that they are correct.

Once this amount of achievable threshold has been established for a particular Air Force application, it remains to find out what actual circuitry will provide this extension of capability, and the costs involved.

Alfred S. Kobos

ALFRED S. KOBOS
Project Engineer

I. INTRODUCTION

The objective of the investigations reported here is to develop a thorough understanding of the benefits which can be achieved in the demodulation of angle modulated (PM and FM) signals, through the use of estimators based on the statistical criteria of minimization of the mean-square error or maximum likelihood estimation.

With angle modulation the modulated transmitted signal is given by

$$\begin{aligned} e_T(t) &= Re\left\{E_0 e^{j(\omega_0 t + \theta[a(\cdot),t])}\right\} \\ &= E_0 \cos\left(\omega_0 t + \theta[a(\cdot),t]\right) \end{aligned} \tag{1}$$

where

$Re\{\ \}$ designates the real part of the expression in the brackets.

E_0 the carrier amplitude.

ω_0 the carrier radian frequency

$\theta[a(\cdot),t]$ the phase angle at time t due to the modulation signal $a(\cdot)$.

The reason that the phase angle is written $\theta[a(\cdot),t]$ and not $\theta[a(t)]$ is that with certain forms of modulation the value of $\theta[\cdot,\cdot]$ at time t is determined not only by the value of $a(\tau)$ at time $\tau = t$, but rather by the entire function $a(\cdot)$. For example, with FM $\theta[a(\cdot),t] = \beta\int_0^t a(\tau)d\tau + \theta_0$. Since in general

$$Re\left\{e^{j(\alpha+\gamma)}\right\} \neq Re\left\{e^{j\alpha}\right\} + Re\left\{e^{j\gamma}\right\}$$

angle modulation is inherently a nonlinear operation even when $\theta[a(\cdot),t]$ is a linear functional. The nonlinear nature of the modulation is directly responsible for such salient characteristics of angle modulated communication systems as the improvement factor (greater noise immunity obtainable with PM or FM than with a linear modulation system, e.g., AM), greater bandwidth occupancy and the threshold effect. This nonlinear characteristic also makes the analytical treatment of these systems very difficult.

The demand for more communication channels has in recent years led to a progressive narrowing of the channel bandwidth assigned to FM voice transmission from 100 kHz to 50 kHz to 25 kHz. In part, these narrower assignments have been made possible by the development of better frequency control systems; however, the overriding consideration has been the demand for more channels. With linear modulation systems where the RF bandwidth of the transmitted signal is independent of the degree of modulation,* the receiver bandwidth can be reduced to a certain point without impairment of performance as better frequency control of the transmitter and receiver becomes possible. This reduced receiver bandwidth may actually result in improved performance because of the decreased amount of noise accepted by the receiver. With double sideband AM, and a 4 kHz audio modulating signal and a corresponding 8 kHz required RF bandwidth, assuming precise tuning, it is obvious that the 25 kHz channel spacing is not dictated by consideration of the required RF bandwidth. On the other hand, with FM modulation the required bandwidth B is approximately given by Carson's rule, $B = 2(\mu+1)W$ so that for a 25 kHz channel spacing and modulation bandwidth $W = 4$ kHz, the modulation index μ, could at most equal

$$\mu = \frac{B}{2W} - 1 = \frac{25}{2\times 4} - 1 = 2.125.$$

With this restriction on the bandwidth of the transmitted signal, the performance may be degraded noticeably compared to that obtainable with wider bandwidth FM systems. As is well known, the above-threshold signal-to-noise ratio performance in a random noise environment will vary as the square of the modulation index. In many cases, a 6 db reduction in output signal-to-noise ratio may be tolerated in order to gain a two-fold gain in the number of channels. Even with a modulation index as low as 2, the FM system will still have a signal-to-noise ratio characteristic superior to AM. However, especially in vehicular applications, the increased susceptibility of FM systems to impulse noise as the bandwidth is reduced may be a far

---

*If these systems are modulated beyond their normal operating range, overmodulated, nonlinear operation will occur, more RF bandwidth will be occupied and the demodulated signal will be distorted.

more important factor than consideration of thermal noise. In fact, one of the major reasons why FM has been employed in the mobile service is because wideband FM systems have been found to be less susceptible to impulsive type interference than AM systems. However, FM systems lose this advantage when small values of modulation index are employed. Consider an FM system with a 4 kHz modulating signal. The instantaneous frequency is

$$\omega(t) = \omega_c + \Delta\omega \cos \omega_m t$$

and the resulting phase is

$$\phi(t) = \omega_c t + \frac{\Delta\omega}{\omega_m} \sin \omega_m t$$

$$= \omega_c t + \mu \sin \omega_m t$$

Thus, μ is numerically equal to the transmitted peak phase deviation (in radians) due to modulation. For very large values of μ, the peak phase excursions can be many radians, while for a μ of 2, the peak phase deviation will be 2 radians or approximately 114.6 degrees. An interfering impulse can cause at most a $\pm$ 180 degree phase error and therefore will affect the output of a narrow-band (low μ) system much more than a wide-band system where the phase perturbation caused by the impulse will be much less than the phase variation created by the modulation.

From the outset of this program it was anticipated that it would be very difficult to find a single technique capable of determining the limiting performance of angle modulated systems over the desired range of input signal-to-noise ratio and modulation index. Therefore, it was decided to undertake several distinct investigations, each of which is most effective in a restricted domain. It was possible to choose these investigations such that the overlap of these domains resulted in coverage of the entire desired range of parameters.

By analytic means, using asymptotic expansions, it was possible to cover the following cases:

1) The large $(S/N)_i$ region, for all values of the modulation index. The results, derived in Chapter II, are valid from large $(S/N)_i$ down to

and in to the threshold region* for values of the modulation index of most interest.

2) The small $(S/N)_i$ region for all values of the modulation index. These results are derived in Chapter III.

3) The region of small modulation indexes for all values of $(S/N)_i$. These results are reported in Appendix C to Chapter IV.

In addition to these analytic efforts the following investigations were performed with the aid of computers.

A digital computer was used to obtain bounds on the performance attainable with PM for $(S/N)_i$ over the range of -15 to +30 db for small to moderate values of the modulation index. This range of $(S/N)_i$ includes the entire threshold region where it is very difficult to obtain results by analytical techniques. This investigation is described in Chapter IV.

A hybrid computer (partly analog and partly digital) was used to determine the performance obtainable by realizable demodulators, the design of which is suggested by application of statistical optimization criteria. The threshold behavior of several phase-locked loop designs was explored for FM systems with bandwidth expansion factors as large as 82. Chapter V contains a detailed description of these investigations.

The modulation signals used were obtained from random processes with statistics chosen by considering both applications and facility of carrying out the investigations. An appropriate statistical description of the modulation process if of course a prerequisite to the application of statistical estimation techniques. Furthermore, it is known that in order to carry any information, in the technical sense, the modulating signals must be generated by random processes. For the purpose of analysis it is convenient to assume that the modulating signal is a sample function generated by a random

---

*The intrinsic input signal-to-noise ratio $(S/N)_i$ is defined as the ratio of RF signal power to RF noise power in the bandwidth of the modulation

$(S/N)_i = \frac{E^2/2}{N_0 W} = \frac{E_c^2}{2N_0 W}$ where E_c . . . carrier amplitude

N_0 . . . noise power density

W . . . bandwidth of modulation

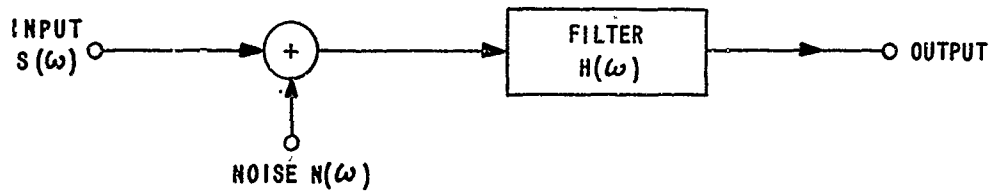
Gaussian process. There exists considerable evidence that results obtained by analysis using Gaussian modulation, or by testing with Gaussian noise loading, can be used to predict the performance obtained with other types of modulation, such as speech or data. It is also noted that in certain applications of great practical importance, e.g., in frequency division multiplex systems, where many independent modulation signals are multiplexed, the statistics of the resulting modulation approach that of a Gaussian process.

In the analytical treatment reported in Chapters II and III, random Gaussian modulation was used. For the computer investigation, reported in Chapter IV, a uniformly distributed modulation process was used. The reason for this choice was that the uniform distribution, in contrast to the Gaussian distribution, does not have "tails". The absence of tails of the uniform distribution results in a more economical computer investigation.

In the hybrid computer simulation, both filtered Gaussian noise and sinusoidal modulating signals were used. The latter were used primarily for comparison with results obtained with tests using the traditional sinusoidal test tones. However, by this means it was also verified that the results obtained with Gaussian modulation can, with caution, be used to predict the performance obtained when other modulating signals (e.g., sinusoids) are employed. For example, the threshold observed with a phase-locked loop, PLL, occurs at a value of $(S/N)_i$ which depends on the degree of modulation present; however, when the rms frequency deviation was kept constant, the threshold occurred at the same value of $(S/N)_i$ with both Gaussian and sinusoidal modulating signals. (This observation was made at a relatively large value of the modulation index and may not hold for small values of the index, although it is obviously applicable in the limit as the index approaches zero.)

In this report we are concerned with determining the performance attainable with angle modulated signals when certain criteria of statistical estimation are utilized to determine the receiver output. The performance of a communication system is ultimately judged in terms of satisfying a user. Because the feedback of user's reaction is a slow process which often yields only qualitative or grossly quantized information (e.g., the system is or is not satisfactory), it is necessary to develop more quantitative measures of

the performance of communication systems. In the case of digital systems the probability of error is usually used as the criterion while with analog systems the output signal-to-noise ratio $(S/N)_o$ is often used. In many applications it is possible to relate the output signal-to-noise ratio to the probability of error when digital data are sent over the communication system. This should not be construed to imply that the output signal-to-noise ratio suffices to specify the performance of a communication system. As a matter of fact, in order for specification of $(S/N)_o$ to be of much use, most of the other parameters of the system, such as bandwidth, noise, and signal spectra, etc., must be known. However, for a given system these parameters may be either fixed or their dependence on $(S/N)_o$ may be known. Under these conditions knowledge of $(S/N)_o$ is useful in determining system performance. Although the concept of output signal-to-noise ratio is widely used, it is essential that it be precisely defined in order that it may reasonably be used as a quantitative measure of the performance of a communication system.



In a linear system, such as sketched above, the output signal-to-noise ratio can be defined as $(S/N)_o = \frac{\text{output power due to signal input}}{\text{output power due to noise input}}$ and the $(S/N)_o$ can be maximized by an appropriate choice of the filter transfer function $H(\omega)$. Unless the signal and noise spectra are identical one finds in this case that maximization of the output signal-to-noise ratio leads to a very narrow band filter which passes just that band of frequencies where the ratio of the signal spectral density $S(\omega)$ to the noise spectral density $N(\omega)$ is greatest. In most cases the output of such a system would be useless for the intended transmission of information. Consequently, a system specification must always require that some other criterion in addition to output signal-to-noise ratio, as defined above, be met. Very often this is stated

in terms of the required frequency response, e.g., the response shall be flat within some tolerance over a certain band of frequencies.

Assuming that the system has been specified so as to be satisfactory in the absence of noise, one is certainly interested in the manner in which its performance degrades as noise is added to the receiver input. If the system is linear it now makes sense to consider the output signal-to-noise ratio as a measure of this degradation. In this case the output signal-to-noise ratio is simply the output power due to the signal input divided by output due to the noise input.

Most communication systems exhibit a quasi linear range where the total output can be considered as consisting of the sum of the output due to signal plus the output due to noise. If one wishes to measure the output signal-to-noise ratio in this region one must supply an input signal which is sufficiently strong so that operation is above threshold. While it makes sense to define the output signal-to-noise ratio as the output power due to the signal divided by the output power due to noise, one certainly cannot measure the output power due to the noise by removing the input signal. If this were done the system would drop below threshold and invalidate the measurement. In some cases, a way out of this difficulty is to remove the modulation while maintaining the carrier. A safer technique is to measure the correlation between the modulation and the noisy receiver output. One now assumes that the system acts in a quasi linear fashion. In other words, it is assumed that while the noise level is influenced by the strength of the carrier that the output noise is independent of the modulation.

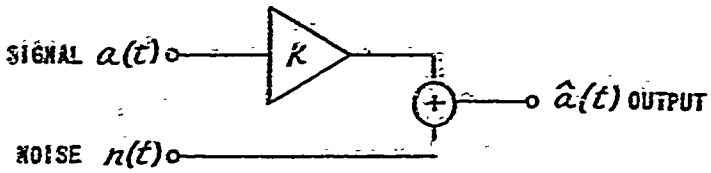
The correlation coefficient $\rho(t)$ between input $a(t)$ and output $\hat{a}(t)$ is defined by

$$\rho(t) = \frac{\langle (a(t)-\langle a(t)\rangle)(\hat{a}(t)-\langle \hat{a}(t)\rangle)\rangle}{\left[\sigma_a^2(t)\,\sigma_{\hat{a}}^2(t)\right]^{1/2}} \tag{2}$$

$$\frac{\langle a(t)\,\hat{a}(t)\rangle}{\left[\langle a^2(t)\rangle\langle \hat{a}^2(t)\rangle\right]^{1/2}} \quad \text{if } \langle a(t)\rangle = \langle \hat{a}(t)\rangle = 0$$

where the angular brackets denote statistical average of the quantity within the brackets and σ_a^2, $\sigma_{\hat{a}}^2$ are the variances of $a(t), \hat{a}(t)$. If $a(t)$ and $\hat{a}(t)$ are stationary processes, then the correlation coefficient is independent of time: $\rho(t) = \rho$. Unless otherwise stated we shall assume that $a(t), \hat{a}(t)$ are stationary and ergodic. The latter assumption permits time averages to be substituted for ensemble averages[1], as was done in the experimental work reported in Chapter V. We note that the correlation coefficient remains unchanged in the presence of linear amplification. The relationship between correlation coefficient and signal-to-noise ratio is easily established for independent additive noise with

$$\langle a(t) \rangle = \langle \hat{a}(t) \rangle = 0 \,, \quad \langle a(t) n(t) \rangle = 0 \,.$$



$$\hat{a}(t) = K a(t) + n(t)$$

$$\rho^2 = \frac{\langle a(t)\,\hat{a}(t) \rangle^2}{\sigma_a^2\,\sigma_{\hat{a}}^2} = \frac{\langle a(t)(K a(t) + n(t)) \rangle^2}{\sigma_a^2 (K^2 \sigma_a^2 + \sigma_n^2)}$$

$$= \frac{K^2 \sigma_a^4}{\sigma_a^2 (K^2 \sigma_a^2 + \sigma_n^2)}$$

$$= \frac{K^2 \sigma_a^2}{K^2 \sigma_a^2 + \sigma_n^2}$$

$$(S/N)_o = \frac{K^2 \sigma_a^2}{\sigma_n^2} = \frac{\rho^2}{1 - \rho^2} \tag{3}$$

For stationary processes it is always possible to experimentally evaluate the correlation coefficient between the modulation and its estimate at the output of any system, linear or otherwise. The quantity $\Lambda_o = \frac{\rho^2}{1-\rho^2}$ which reduces to the output signal-to-noise ratio for the case where only noise is added to the signal is used as a measure of system performance in parts of Chapters III and IV.

It is not difficult to show (via Appendix C of Chapter IV) that $\Lambda_o = \frac{\rho^2}{1-\rho^2} = \frac{\sigma_a^2}{MSE} - 1$ provided that the level of the output is adjusted to minimize $\langle (a - \hat{a})^2 \rangle \equiv MSE$ which would of course be the case with minimum mean-

square error estimation. The relationship of Λ_0 vs $\frac{\sigma_a^2}{MMSE}$ is shown in Figure I-1.

In all of the investigations reported in this report we have assumed that the received signal is perturbed by additive white Gaussian noise of spectral power density N_0 watts/Hz over the RF bandwidth B of the receiver. Many of our results are reported in terms of the intrinsic input signal-to-noise ratio $(S/N)_i \equiv \frac{E_0^2}{2N_0W}$

where E_0 . . . carrier amplitude, volts

N_0 . . . noise spectral density watts/Hz.

W . . . bandwidth of the modulation, Hz.

Both B and W are one-sided bandwidths (positive frequencies only) and N_0 is a one-sided power spectral density. The noise in the RF bandwidth is then BN_0 and the signal-to-noise ratio in the RF bandwidth is

$$(S/N)_{RF} = \frac{E_0^2}{2N_0B} = \frac{E_0^2}{2N_0W}\,\frac{W}{B} = \frac{1}{M}\,(S/N)_i \qquad (4)$$

where $M = \frac{B}{W}$ is the bandwidth expansion factor.

As the RF input signal-to-noise ratio $(S/N)_i$ becomes very large, the minimum MSE goes to zero and $\Lambda_0 = \frac{\sigma_a^2}{MSE} - 1 \sim \frac{\sigma_a^2}{MSE}$ This is confirmed by the experimental data presented in Chapter IV where plots of Λ_0 and $\frac{\sigma_a^2}{MSE}$ vs $(S/N)_i$ with all quantities expressed in db, have the same asymptote at large $(S/N)_i$.

As the RF input signal-to-noise ratio approaches zero the output of the receiver becomes independent of the modulation of the RF input signal. Under these conditions the minimum mean-square error in estimating the modulation approaches the power of the modulation, MMSE $\sim \sigma_a^2$. Consequently, $\Lambda_0 = \frac{\sigma_a^2}{MSE} - 1$ goes to zero as $(S/N)_i$ approaches zero. This is also confirmed by the results of Chapters III and IV where $\frac{\sigma_a}{MSE} \longrightarrow 0\,db$ and $\Lambda_0 \longrightarrow -\infty\,db$ as $(S/N)_i \longrightarrow -\infty\,db$.

The observation that the MMSE $\sim \sigma_a^2$ as the input signal-to-noise ratio goes to zero implies that when the input signal-to-noise ratio is very small the best estimate, in the mean-square error sense, goes to zero. This observation should not be too startling. As a matter of fact, very

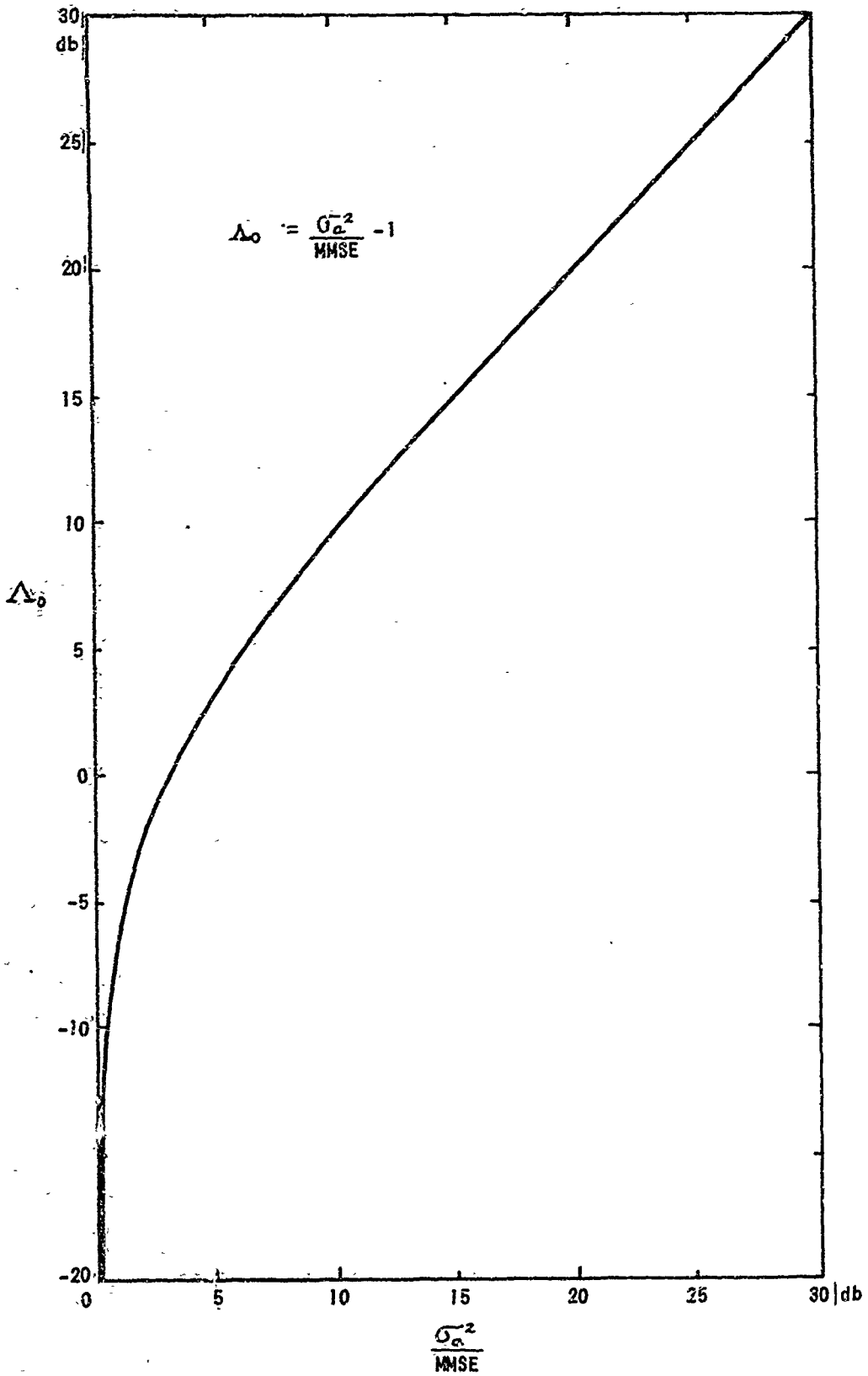

30
db
25
20
15
10
5
0
-5
-10
-20
$\Lambda_o = \frac{\sigma_a^2}{MMSE} - 1$
Λ_o
0
5
10
15
20
25
30 db
$\frac{\sigma_a^2}{MMSE}$

Figure I-1

similar approaches have been in practical use for many years, e.g., noise limiter circuits which reduce the output of a receiver to zero when the input noise is so large that a usable output signal cannot be obtained and squelch circuits which disable the receiver unless a sufficiently strong carrier is received.

A major objective of this contract is to evaluate the performance attainable through the use of the criterion of minimum mean-square error estimation; in this case use the ratio of

$\frac{\text{modulation power}}{\text{mean-square error}} = \frac{\langle a^2 \rangle}{MSE} = \frac{\sigma_a^2}{MSE}$ for $\langle a \rangle = 0$, which is often proposed as an alternate definition of output signal-to-noise ratio, is particularly appropriate as a measure of performance.

However, the quantity $\frac{\sigma_a^2}{MSE}$ also plays a fundamental role in information theory where it is closely related to the rate-distortion function.[1, 2, 3] Shannon has shown that in order to transmit an analog signal through a communication channel with a given mean-square error between output and input, a minimum information rate $R[a(\cdot), MSE]$ must be provided. (This notation indicates that the information rate $R[\cdot, \cdot]$ is a function of both the modulation process $a(\cdot)$ and the permissible mean-square error, MSE). For the case where $a(\cdot)$ is a white Gaussian process limited to bandwidth W Hz

$$R[a(\cdot), MSE] = W \log_2 \frac{\sigma_a^2}{MSE} \quad \frac{\text{bits}}{\text{sec}} \tag{5}$$

It is emphasized that equation (5) does not apply to other than white Gaussian processes. However, the relationship

$$W \log_2 \frac{P_1(a)}{MSE} \leq [R\ a(\ ), MSE] \leq W \log_2 \frac{\sigma_a^2}{MSE} \cdot \tag{6}$$

where $P_1(a)$ is the entropy power* of $a(\cdot)$, holds, for any source limited to bandwidth W Hz. For white Gaussianly distributed $a(\cdot)$, $P_1(a) = \sigma_a^2$ so that (5) applies.

*The entropy power of a random process is defined as the power of a white Gaussian process having the same bandwidth and entropy as the process under consideration. [1]

In order to transmit information at the rate $\mathcal{R}[a(\cdot), MSE]$ the channel capacity, C, must be equal to or greater than this rate. In general it is not known how to transmit a(.) over a given channel so as to obtain a desired MSE when that channel has the minimum required capcity $C = \mathcal{R}[a(\cdot), MSE]$.

However, in any communication system with given channel conditions a certain MSE will result and this MSE can usually be decreased by improving the channel conditions. This will, however, require a channel of capacity greater than $\mathcal{R}[a(\cdot), MSE]$. The difference between the actual capacity and the minimum required capacity, or the actual $(S/N)_i$ and the minimum theoretically required $(S/N)_i$ are indications of how efficiently the channel is being utilized. This difference is in general attributable to several causes:

The modulation (coding) used may not be optimum for use with the available channel. For example, with angle modulation the spectrum of the transmitted signal does not in general fill the channel uniformly; this can be interpreted as a manifestation of suboptimal coding. The demodulation (decoding) used may not be optimum; the optimum demodulation process is unknown for arbitrary forms of modulation processes.

In the investigations reported here the modulation was constrained to be angle modulation (PM or FM) and it was attempted to determine the optimum performance attainable under this constraint with the intent of comparing the performance actually attained by known techniques for the demodulation of these signals against the optimum. Our investigations show that if the modulation index is held constant and the channel capacity is varied by changing the input signal-to-noise ratio $(S/N)_i$, the available channel capacity is used most efficiently when operation is near or somewhat below the threshold region. From a practical point of view this is a very difficult region in which to operate because small changes in $(S/N)_i$ result in large changes in output signal-to-noise ratio. If $(S/N)_i$ is increased, so that operation is above threshold, the ratio $\frac{\sigma_a^2}{MSE}$ obtained increases, but at a lesser rate than theoretically made possible by the increase in channel capacity. As a matter of fact, with angle modulation above threshold $\frac{\sigma_a^2}{MSE}$

increases at most linearly* with $(S/N)_i$ whereas the limiting $\frac{\sigma_a^2}{MSE}$ increases as $(S/N)_i^M$,

where M is the bandwidth expansion factor

$$M = \frac{W_{CH}}{W} = \frac{\text{RF channel bandwidth}}{\text{Bandwidth of the modulation}} \tag{7}$$

It is therefore not "efficient" to operate such systems far above threshold. In order to obtain efficient operation without being unduly sensitive to fluctuations in $(S/N)_i$, the modulation index should be adjusted such that operation is not too far above the threshold region. This may, of course, not be practical, for instance when the available $(S/N)_i$ is large the modulation index would have to be increased to a value which would result in a transmitted spectrum which exceeds the assigned RF bandwidth. Efficient utilization of the transmitted power under these conditions could, however, be maintained by increasing the modulation bandwidth.

The capacity of a channel perturbed by additive Gaussian noise is given by the famous formula[1]

$$C = \ln \left(1 + \frac{S}{N_0 W_{CH}}\right)^{W_{CH}} \text{ nits/sec.} \tag{8}$$

which takes the limiting form

$$\lim_{W_{CH} \to \infty} C = e^{S/N_0}. \tag{9}$$

---

*The rate of increase of $\frac{\sigma_a^2}{MSE}$ vs $(S/N)_i$ expressed in db above threshold depends upon the spectrum $S_a(\omega)$ of the modulation. A unity slope is obtained for a strictly bandlimited spectrum, while a first order Butterworth spectrum results in a slope of .25 and a third order Butterworth spectrum in a slope of approximately 0.6. These results are due to a widening of the frequency response of the receiver as $(S/N)_i$ increases so as to minimize the total MSE consisting of a distortion and a noise component. If the receiver response is kept fixed, then the output noise component above threshold always decreases 1 db for each db of increase in $(S/N)_i$ [4].

Assuming that the modulation process is white Gaussian with bandwidth W and combining Equations (5) - (9), one obtains

$$\frac{\sigma_a^2}{MSE} = \left[1 + 1/M(S/N)_i\right]^M \tag{10}$$

$$\lim_{M\to\infty} \frac{\sigma_a^2}{MSE} = e^{(S/N)_i} \tag{11}$$

as bounds on the performance attainable with any combination of modulation and receiver. These expressions are plotted in Figure I-2 together with experimentally derived bounds on $\frac{\sigma_a^2}{MSE}$ for PM derived in Chapter IV. The interpretation of these bounds is that no communication system can exceed the performance (plot above the solid curves) described by equations (10), (11), and that a PM system cannot exceed the performance depicted by the dashed curves. The experimental curves were obtained by use of a digital computer, as explained in detail in Chapter IV. In that chapter it is pointed out that there is some difficulty in defining the RF bandwidth for randomly angle modulated signals. Bandwidth expansion factors of 10 and 20 are considered to be "generous" for the values of $\beta = 4$ and 8, used in the curves shown. It will be observed from Figure I-2 that a change by even a factor of 2 in the value of M shifts the value of $(S/N)_i$ at the level of $\frac{\sigma_a^2}{MSE}$ corresponding to the threshold by only a very small amount. It is clear that the PM bounds closely parallel the general bounds derived from information theory below threshold, but deviate rapidly above threshold.

For $(S/N)_i$ such that operation is at least a few db above threshold, the performance of angle modulated systems can be easily computed by linear analysis (viz Appendix C of Chapter IV). Furthermore, it is found that linear optimization techniques employed above threshold lead to the same performance for all demodulators, e.g., the above threshold performance of a properly designed receiver using a discriminator will be the same as that for a phase-locked loop. However, different demodulation techniques can result in drastically different values of the $(S/N)_i$ at which threshold occurs. The normal operating range for angle modulated systems is restricted to above threshold operation, in fact, for the comparions of various demodulators the value of $(S/N)_i$ required for above threshold

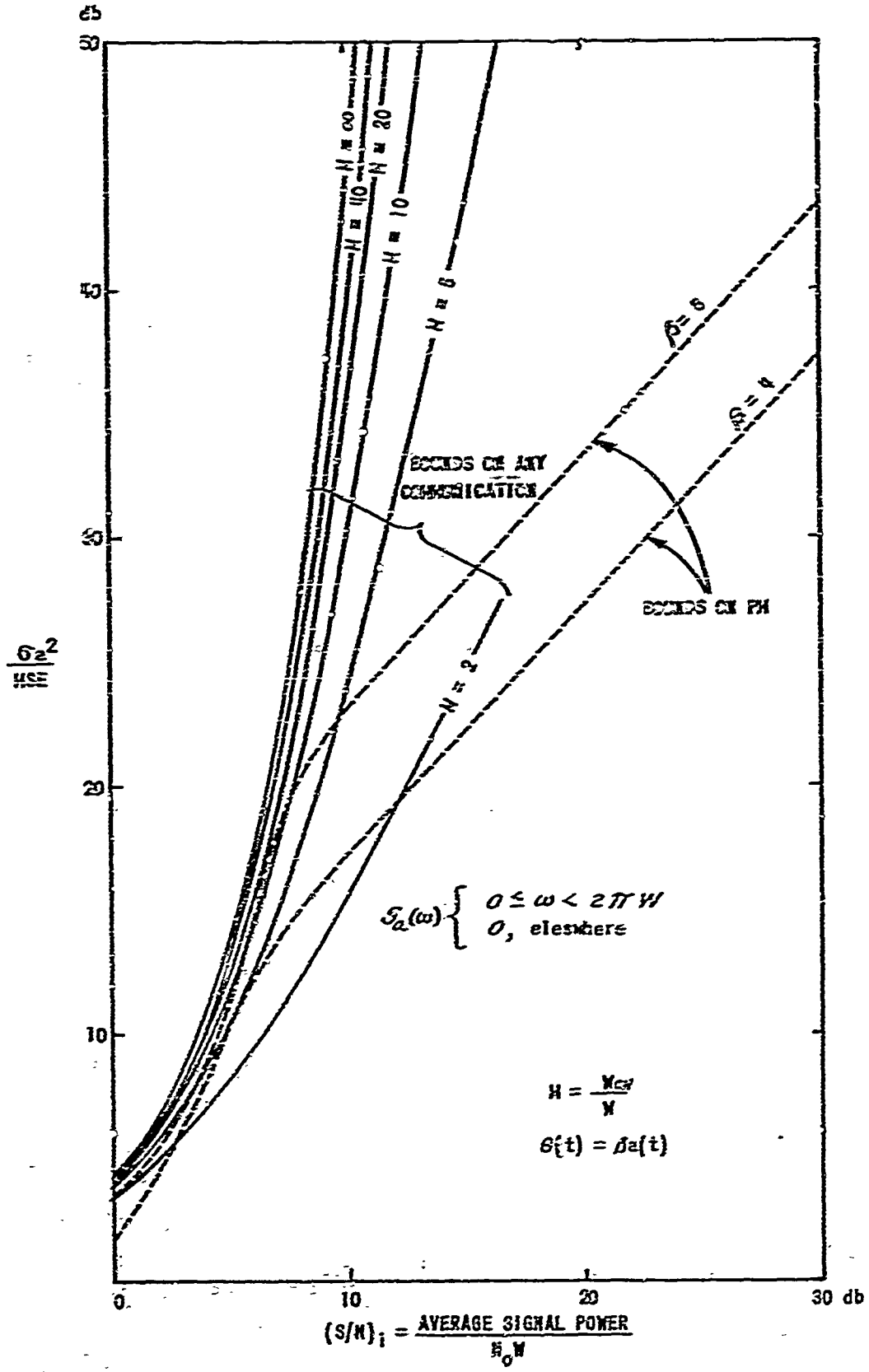


Figure 2-2 INFORMATION THEORY BOUNDS ON PERFORMANCE, BOUNDS ON PERFORMANCE OF PM

operation is often considered to be the parameter of greatest practical importance. The value of $(S/N)_i$ required for above threshold operation will be a function of (1) the type of receiver, (2) the degree of modulation, and (3) the spectrum of the modulation.

The value of $(S/N)_i$ required for above threshold operation always increases with increasing modulation index (increasing RF bandwidth). In the older systems, using limiters and discriminators, occurrence of the threshold is determined primarily by the signal-to-noise ratio in the RF bandwidth. Consequently, in these systems the minimum intrinsic signal-to-noise ratio $(S/N)_i$ required for above threshold operation has to increase approximately linearly with increasing bandwidth expansion factor. With more recent designs using frequency modulation with feedback [5] (FMFB), or phase-locked loops, [6, 7] (PLL), the intrinsic signal-to-noise ratio $(S/N)_i$, increases less rapidly with increasing RF bandwidth.

The minimum value at which threshold can occur can be determined as follows. Consider for example, FM.

The transmitted FM signal is given by

$$e_T(t) = E_c \cos\left(\omega_c t + \beta \int_0^t a(u)\, du\right) \tag{12}$$

The instantaneous radian frequency deviation is then β a(t) rad/sec. With sinusoidal modulation a(t) = sin $2\pi W t$ so that $\Delta f_{max} = \frac{\beta}{2\pi}$, and the modulation index $\mu \equiv \frac{\Delta f_{max}}{f_m} = \frac{\beta}{2\pi W}$.

It has been found that for this case negligible distortion occurs if an RF bandwidth given by Carson's rule

$$B = 2f_m(\mu+1) = 2(\Delta f_{max} + f_m) \tag{13}$$

is used. Certainly, the amount of distortion due to use of an RF bandwidth limited to this value, is negligible compared to the output noise when operation is near threshold. If the receiver has a flat response from 0-W Hz the

output signal-to-noise ratio above threshold is found to be

$$(S/N)_{out} = \tfrac{3}{2}\mu^2 (S/N)_i = \tfrac{3}{2}\left(\tfrac{M}{2}-1\right)^2 (S/N)_i \qquad (14)$$

If the signal distortion component is negligible we can equate $(S/N)_{out}$ to $\frac{\sigma_a^2}{MSE}$.

Now consider a Gaussian random modulation function having a flat spectrum over the band 0-W Hz and operating with the same value of β. In order to minimize the MSE the receiver must have a flat response over the band 0-W Hz when the input signal-to-noise ratio is sufficiently large.[8] With a fixed receiver operating above threshold the output noise is independent of the modulation, provided that the RF bandwidth B is sufficiently large. Therefore, in order to obtain the same output signal-to-noise ratio for a given $(S/N)_i$ the same modulating power must be used, e.g., the standard deviation of the modulation $\sigma_a = \frac{1}{\sqrt{2}}$. In the case of sine wave modulation the peak radian frequency deviation was β and the modulation frequency was W. For the case of Gaussian random modulation it is not possible to uniquely specify either a peak frequency deviation or a modulation frequency. Nevertheless, a bandwidth adequate for any specified distortion can be found. In order to make the results obtained directly comparable with the sinusoidal case it would be desirable to define the peak deviation as $\beta\sqrt{2}\,\sigma_a = \beta$ and the modulation frequency as W. The probability that this "peak deviation" is exceeded is approximately .16, the modulation frequency is clearly conservative since the modulating spectrum is limited to 0-W Hz, e.g., the average frequency is $\frac{W}{2}$ Hz. It therefore appears reasonable to use the sinusoidal definition of bandwidth for the purpose of computing the bounds on the minimum $(S/N)_i$ required for above threshold performance. (It will turn out that the results obtained are not strongly dependent on the definition of bandwidth used.)

In Figure I-3 the information theory bounds, Equations (10), (11), and the above threshold FM performance, Equation (14), are plotted. It is seen that for sufficiently small $(S/N)_i$ the linear analysis predicts values of $\frac{\sigma_a^2}{MSE}$ greater than permitted by information theory. This means that the linear analysis is not applicable in this region.

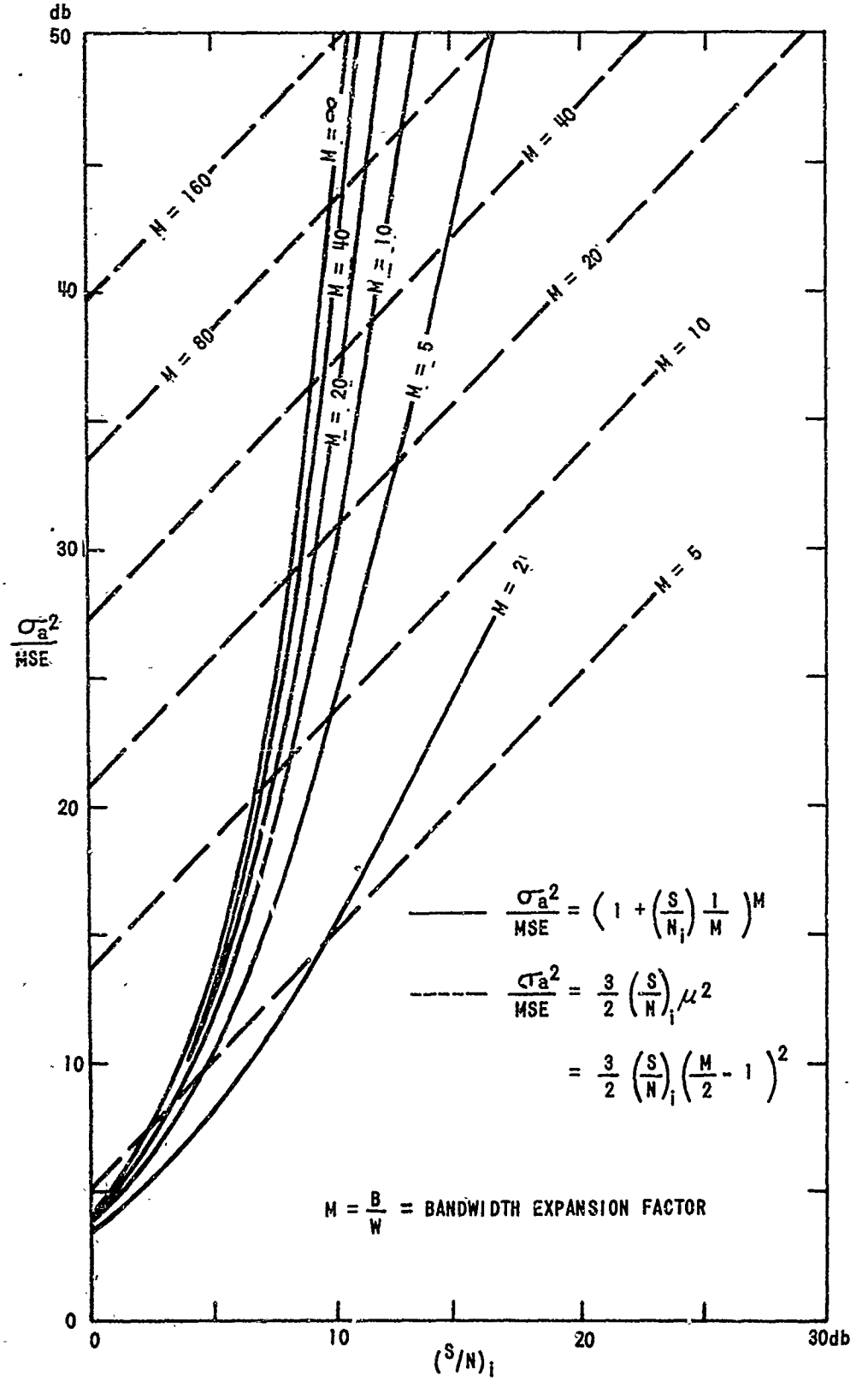


Figure I-3 INFORMATION THEORY BOUNDS ON PERFORMANCE AND ABOVE THRESHOLD FM PERFORMANCE

Threshold may be defined as the point below which the performance deviates by some appreciable amount from that predicted by the linear analysis. We therefore conclude that threshold must occur at higher values of $(S/N)_i$ than given by the intersection of the corresponding (same value of M) straight and curved lines shown in Figure I-3. The values of $(S/N)_i$ at the intersections can be used to obtain a bound on the attainable threshold with FM versus bandwidth expansion factor. The curve obtained in this manner is shown as curve A of Figure I-4, which illustrates the dependence of the value of $(S/N)_i$ at threshold on bandwidth expansion factor for various FM demodulation systems, as well as the bounds derived by application of information theory. For $M > 10$ the slope of curve A is less than 1 db per octave. Therefore, the bounding value of $(S/N)_i$ at threshold is not critically dependent on the value of bandwidth expansion factor used.

Curve B is a plot of results obtained by H. Akima[9], who considered a system where the modulation signal is sampled and "boxcarred" at the Nyquist rate of 2W times per second. The demodulator consists of a bank of contiguous filters each 2W Hz wide, each followed by an amplitude detector and a conventional limiter discriminator. A "greatest of" circuit compares the outputs of the amplitude detectors and selects the filter which is assumed to contain the signal. The output of the corresponding discriminator is added to the center frequency of the selected filter to form the estimate of the frequency of the transmitted signal. Akima's work assumed sinusoidal modulation at frequency W. From Figure I-4 it is observed that Akima's curve closely parallels curve A. This implies that asymptotically with large M frequency modulation uses bandwidth efficiently. The absolute difference between curves A and B is ascribed to:

1) Curve A is a lower bound for FM, i.e., it is only known that FM threshold cannot occur below this curve.

2) Curve B uses only realtime processing up to the output of the discriminators. A comparable restriction does not apply to curve A.

3) It has not been proven that the curve derived from Akima's results represents the limiting performance obtainable with FM.

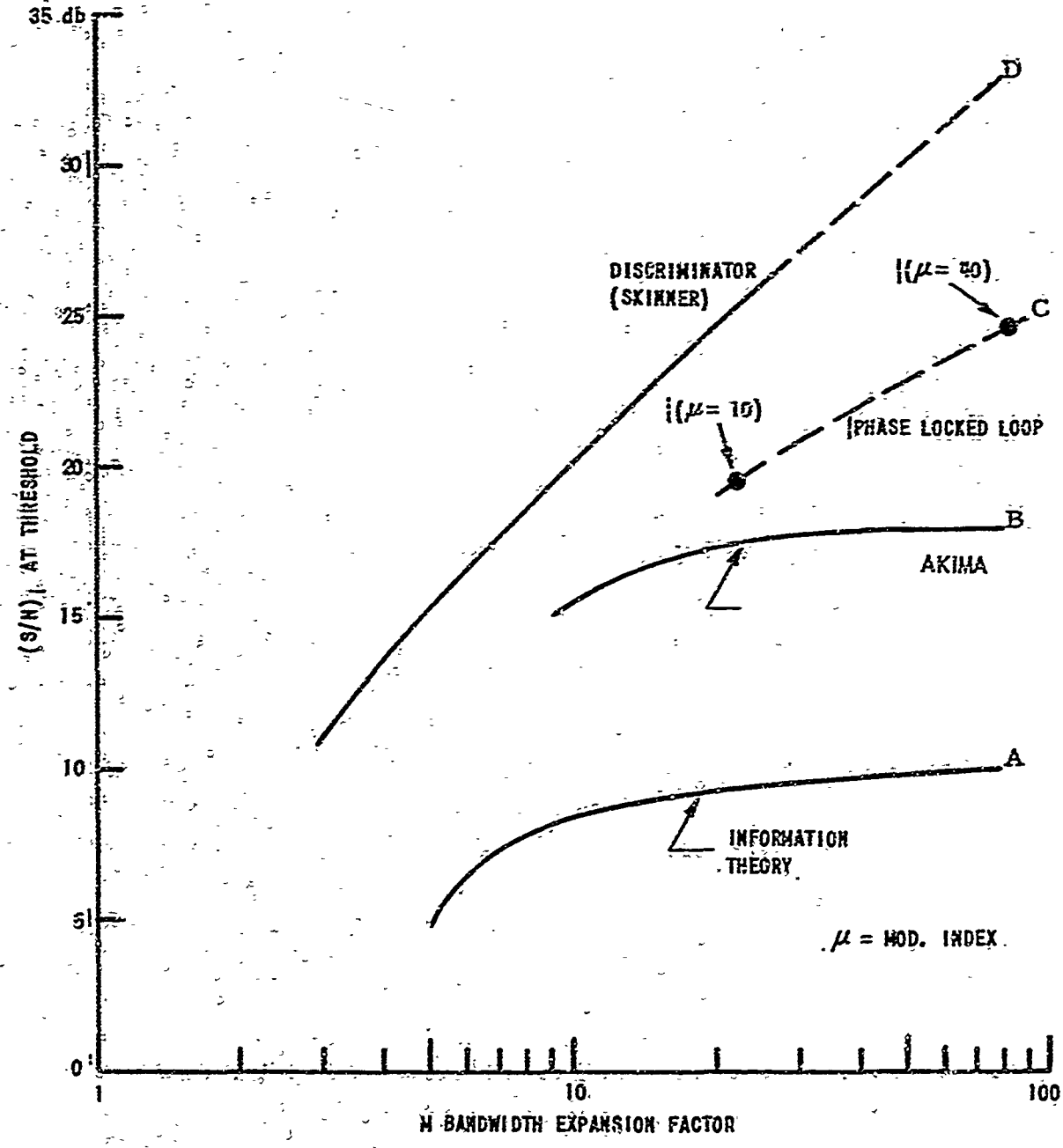


Figure I-4 COMPARISON OF FM DEMODULATOR THRESHOLD PERFORMANCE

Curves C and D represent the threshold performance of realizable FM receivers operated with sinusoidal modulation. Curve D was obtained from Skinner's work[10]; it applies to the use of a conventional discriminator. In this case the threshold occurs at a slowly increasing signal-to-noise ratio as measured in the RF bandwidth.

This is illustrated in the table below:

$$(S/N)_{RF} = (S/N)_i$$

M	5	10	40
M db	7	10	16
$(S/N)_i$ db	15.7	20.6	28.8
$(S/N)_{RF}$ db	8.7	10.6	12.8

Curve C represents results obtained from the simulation of realizable phase-locked loops, as described in detail in Chapter V. The results show that a significant threshold improvement, with respect to a discriminator, can be achieved with conventional PLL systems for the values of modulation index normally encountered. In the PLL systems, the threshold appears to occur when the rms value of the loop phase error reaches approximately 0.5 radians. It might be noted that this value of phase error at threshold was obtained in the PLL simulation and also observed as the filtered phase error at threshold in the Monte Carlo simulation of the bounds on the MMSE and MAP estimation in a PM system. The $(S/N)_i$ at which the threshold occurs will be determined by the design of the particular system. In the case of a PLL, the threshold occurs at a more nearly constant signal-to-noise ratio (approximately 6-7 db) as measured in the RF bandwidth, than is the case with a discriminator. This can be deduced from Figure I-4 where the slope of curve C is seen to be intermediate between the slopes of curves A (or B) and D. As far as the design of phase-locked loops is concerned it must be emphasized that the analytic work on maximum a posteriori estimation leads to phase-locked loops which contain unrealizable components.

It turns out that the transfer functions of the loop filter specified by this analysis are identical with those which one obtains if one replaces the sinusoidal nonlinearity with a linear gain and performs the usual infinite delay Wiener filter design. A physically realizable phase-locked loop must, however, utilize realizable, zero delay components. It has often been suggested that the optimum PLL design is obtainable by use of the realizable, zero delay Wiener filters. As long as the PLL operates in an essentially linear manner, i.e., above threshold, optimum performance is obtainable in this manner, provided the appropriate compensating filter is used at the output of the PLL. However, we do not know of any evidence which guarantees that this procedure will lead to the best performance in the threshold region. On the contrary, we have found that by empirically changing the loop gain and filter transfer function from those specified by the above design procedure, somewhat improved performance can be obtained in the threshold region. It has been reported[11, 12, 13] that replacing the conventional phase detector, which produces the sinsuoidal nonlinearity called for by the theory of MAP estimation, in a phase-locked loop, with some other type of phase detector, can result in improved PLL performance, e.g., threshold extension.

A significant improvement in the demodulation of angle modulated signals may be possible if the application does not require realtime processing. The operation specified by the infinite delay MAP analysis could be closely approximated if the received signal plus noise were available in stored form, as for example by recording the I. F. signal. Demodulation might require computer processing and would entail some delay, at least of the order of the correlation time of the modulation, before the estimate of the modulation would become available. In many applications the delay could be so small that these systems would be entirely practical for situations where "realtime" demodulators are ordinarily thought to be a necessity.

In Chapter II the problem of performing the minimum mean-square error estimation of the modulation of a randomly angle modulated signal is treated analytically. Because of the nonlinear modulation process this is a difficult problem and several assumptions, the validity of which requires further justification, had to be made in order to be able to carry the analysis

through to completion. The results obtained for the PM case were used to compute the curves of Figure I-5. In Figure I-6 the results obtained by the computer simulation reported in Chapter IV, normalized to the same scale as Figure I-5, are shown superposed on these analytical results. (It is worthy to note that these curves are scaled directly from Figure IV-1, which was drawn before the analytic results were available.) These curves differ primarily in the difference of $(S/N)_i$ at threshold. The important point is that the results of Chapter IV are performance bounds which cannot be exceeded, whereas the results of Chapter II are due to specific steps which should be implementable in practice. From the discussion of the limits imposed by information theory it is known that the threshold must move toward higher $(S/N)_i$ as β increases; consequently, the analytically obtained results would for very large β ultimately be untenable. It is believed that this inconsistency is due to one of the above mentioned assumptions.

Chapter III develops an analytic evaluation of the performance obtained by angle modulated systems utilizing minimum mean-square error estimation and operating under conditions of very low $(S/N)_i$. Using $\Lambda_o = \frac{\rho^2}{1-\rho^2}$ as the criterion of performance asymptotic expressions for Λ_o in terms of $(S/N)_i$ and the modulation index are derived. Λ_o is found to vary linearly with $(S/N)_i$ when $(S/N)_i << 1$ for all modulation spectra $S_a(\omega)$. The asymptotic results indicate that the severity of the threshold defined as the ratio of the large $(S/N)_i$ asymptote to the small $(S/N)_i$ asymptote increases exponentially with the square of the modulation index, viz Figure I-7.

The major portion of the difference in the performance indicated by the analytic results, including the digital computer simulation and those obtained in practice, or by the analog computer simulation, is attributed to the use of so-called nonrealizable data processing. This means in practice that the estimate of the modulation $a(\tau)$ at time τ could not become available until sometime thereafter, that is at time $t = \tau + \Delta$. It is important to realize that the amount of delay, Δ, required in order to obtain essentially all of the improvement which can be obtained with infinite delay is of the order of the correlation time of the modulation. For most applications the irreducible delay would be of little consequence.

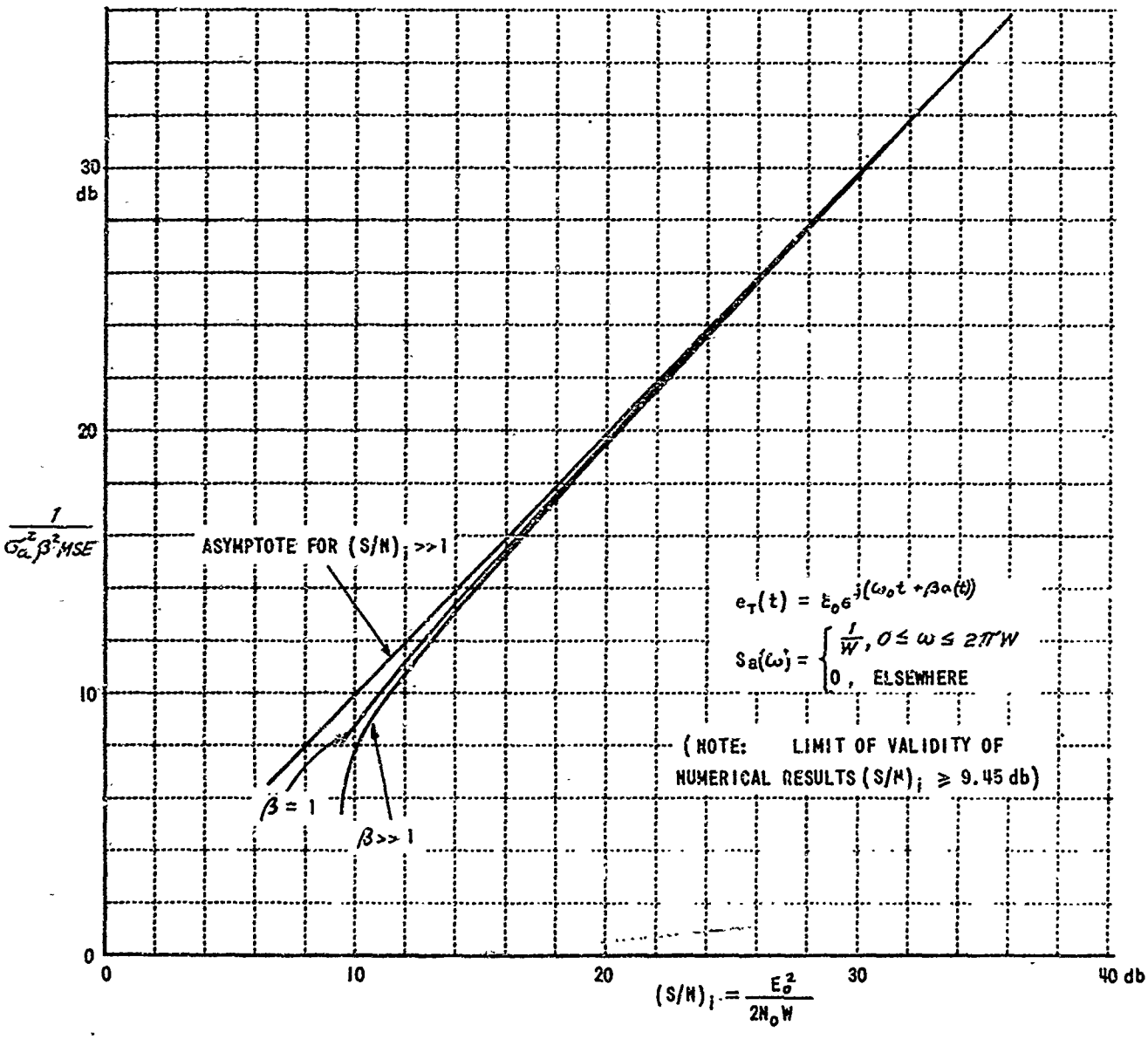


Figure I-5 RESULTS OF ANALYTIC EVALUATION OF MMSE PERFORMANCE OF PM DEMODULATION

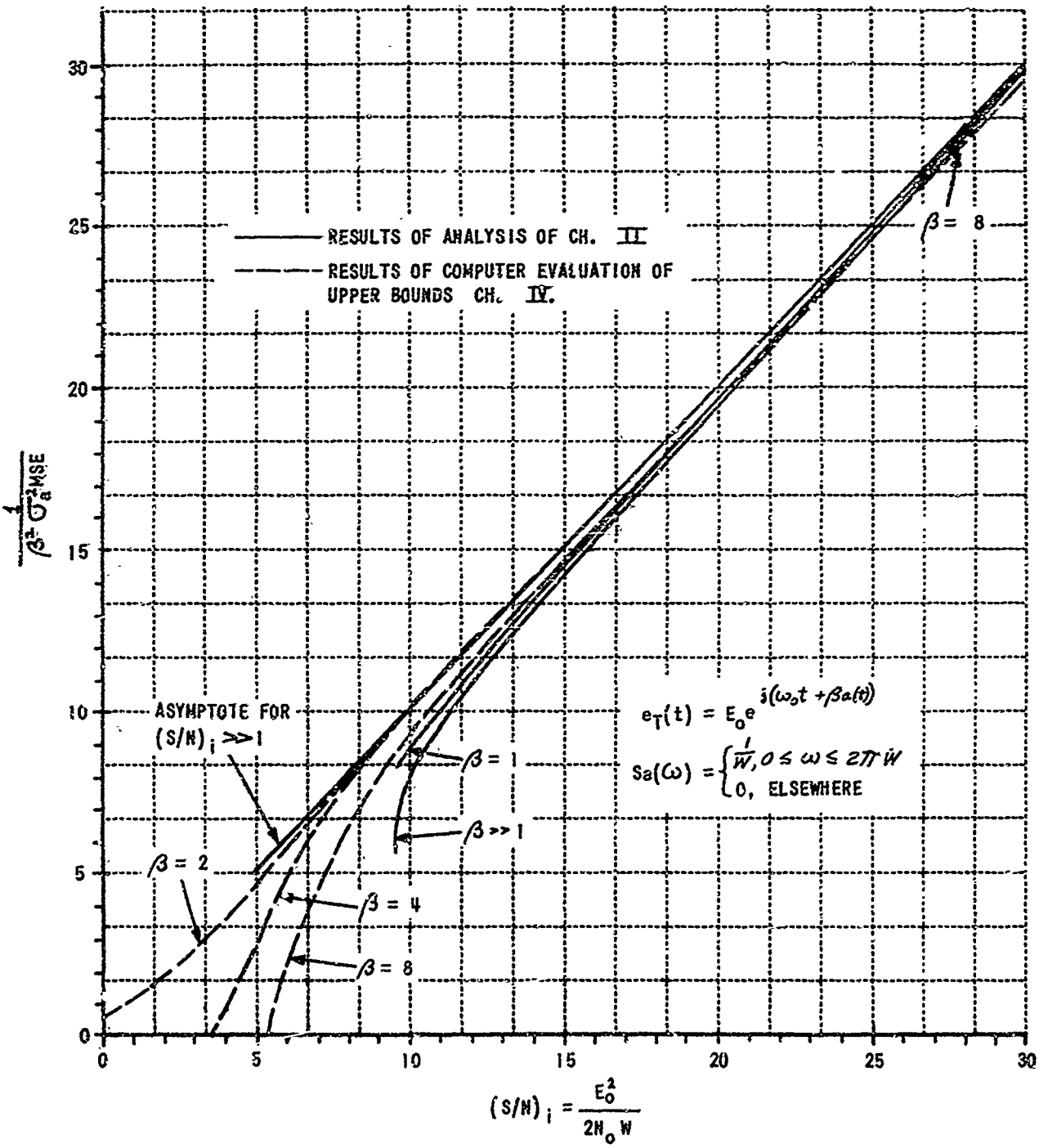


Figure I-6 COMPARISON OF ANALYTIC EVALUATION AND COMPUTER OBTAINED BOUNDS ON THE PERFORMANCE OF PM DEMODULATION

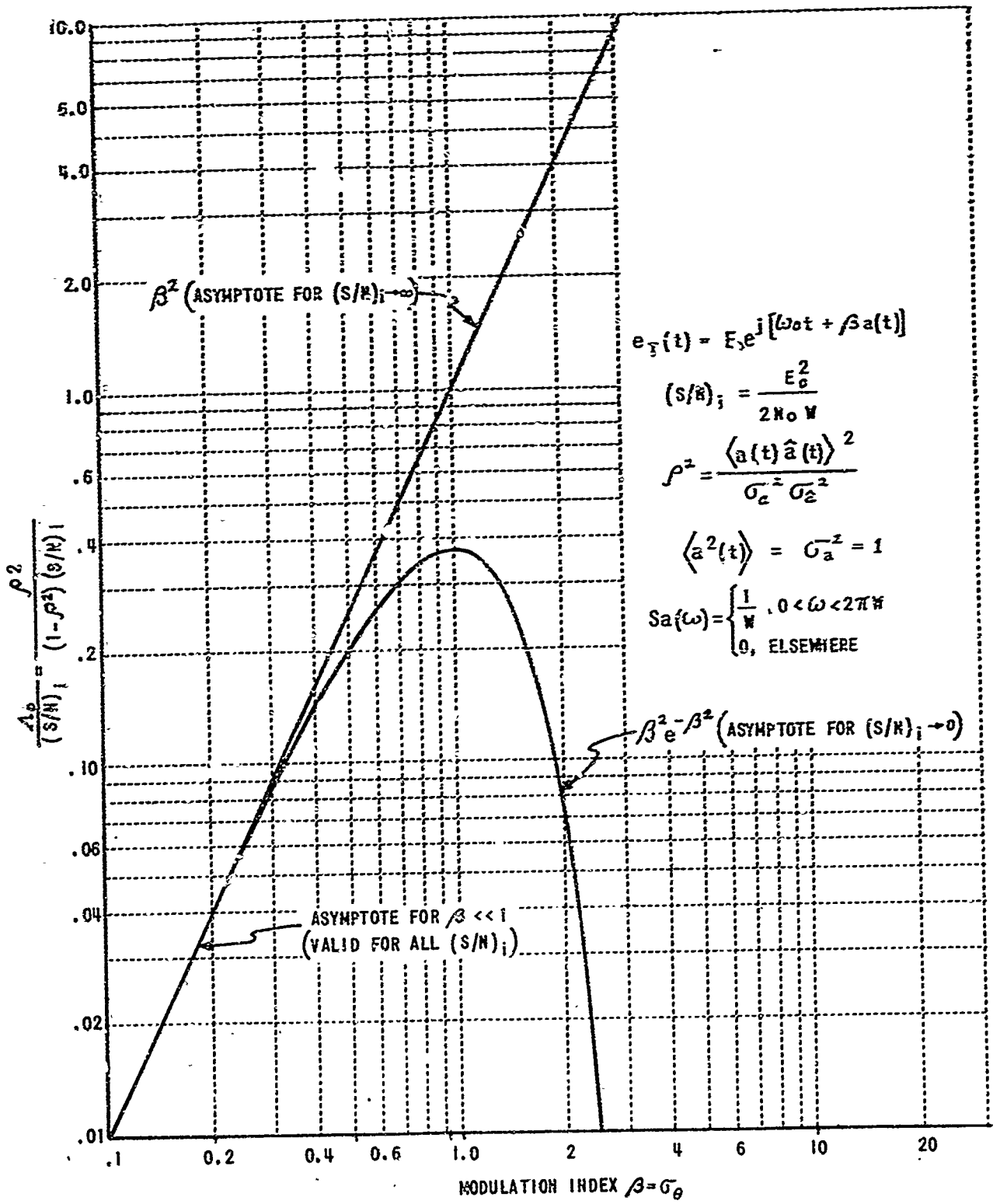


Figure I-7 ASYMPTOTIC PERFORMANCE AT PHASE MODULATION, INFINITE DELAY

The intensive efforts made by many investigators over the past several years notwithstanding, there is no evidence that the limiting performance of angle modulated systems has been attained, and we anticipate that significant improvements in performance will be achieved in the future.

Recommendations

1) The major portion of the difference in the performance theoretically attainable as derived by analysis and digital computer simulation and those currently attained in practice is attributed to the use of so-called nonrealizable data processing (i.e., not realizable without delay). Analysis shows that MAP and MMSE estimation using nonrealtime estimation gives essentially similar results. A technology to implement these estimators should be developed. Implementation would require the use of storage and computers.

2) The configuration of the realizable (zero delay) receiver for angle modulated signals which gives optimum performance in the threshold region is unknown. Efforts aimed at development of improved receivers should be continued.

3) The investigations reported here were concerned with a situation where the only disturbance is additive white Gaussian noise. Real channels are often subject to fading, which results in severe amplitude and phase fluctuations. The analysis of such situations is very difficult and use of channel simulators or experimental investigations have not proven entirely satisfactory. The use of computer simulation techniques for these investigations is recommended.

4) In order to carry the analysis reported in Chapter II through to completion it was necessary to make a number of assumptions, which have not been fully justified. The results obtained predict a performance which asymptotically with large modulation index, exceeds bounds derivable from information theory. This inconsistency should be resolved and the required rigor introduced into the analysis.

5) The digital computer investigation reported in Chapter IV has been constrained by economic considerations to a rather limited range of the modulation index β and use of only uniform distribution of the amplitude of the modulation. For better comparison with the analytic results, this investigation should be expanded to include a wider range of β and Gaussianly distributed modulation processes.

REFERENCES

[1] Shannon, C.E. and Weaver, W. "The Mathematical Theory of Communication", The University of Illinois Press: Urbana, 1949.

[2] Shannon, C.E. "Coding Theorems for a Discrete Source With a Fidelity Criterion", 1959 IRE National Convention Record, Pt. 4, p. 142.

[3] Goblick, T.J. "Theoretical Limitations on the Transmission of Data from Analog Sources", IEEE Transactions on Information Theory, Vol. IT-11, No. 4, October 1965, pp. 558-567.

[4] Lawton, John G. "Investigation of Analog and Digital Communication Systems", Phase 3 Report, RADC-TDR-63-147, ASTIA Document No. AD-407343, May 1963.

[5] Enloe, L.H. "Decreasing the Threshold in FM by Frequency Feedback", Proc. of the IRE, Vol. 50, pp. 18-30, January 1962.

[6] Boardman, C.J. and Van Trees, "Optimum Angle Modulation", IEEE Transactions on Communication Technology, Vol. COM-13, No. 4, pp. 452-468, December 1965.

[7] Viterbi, A.J. "Principles of Coherent Communication", McGraw-Hill, Inc. 1966.

[8] Wiener, N. "Extrapolation, Interpolation and Smoothing of Stationary Time Series", John Wiley, New York City, 1949.

[9] Akima, Hiroshi, "Theoretical Studies on Signal-to-Noise Characteristics of an FM System", IEEE Transactions on Space Electronics and Telemetry, Vol. SET-9, No. 4, December 1963, pp. 101-108.

[10] Skinner, F.J. "Radio Transmission Systems - Theoretical Noise Performance Curves for Frequency Modulation Receivers Operating Below the Breaking Region", Bell Telephone Lab., unpublished memo, File 36690-1, 1 February 1954.

[11] Splitt, Frank G. "Design and Analysis of a Linear Phase-Locked Loop of Wide Dynamic Range", IEEE Transactions on Communication Technology, August 1966.

[12] Robinson, L.M. "Tanlock: A Phase-Lock Loop of Extended Tracking Capability", Proc. of National Winter Conference on Military Electronics, 1962.

[13] Frankle, J. "Threshold Performance of Analog FM Demodulators", RCA Review, December 1966, pp. 521-562.

II. MINIMUM MEAN SQUARE ERROR ESTIMATION OF RANDOM ANGLE MODULATED SIGNALS; WITH EMPHASIS ON THE THRESHOLD REGION

1. Summary

A phase modulated signal of the form

$$e_R(t) = E_o \cos\left(\omega_o t + \beta a(t)\right) + n(t), \; t_o < t < t_N \qquad (1)^*$$

is considered.

We study the problem of determining the MMSE (minimum mean-square error) estimator of the modulating signal $a(t)$ given the received signal $e_R(t)$ in the presence of bandlimited additive Gaussian noise $n(t)$. The method used consists of the following steps:

1) Eliminate the carrier ω_o by considering the in-phase and quadrature components e_1 and e_2 where

$$e_1(t) = E_o \cos \beta a(t) + n_1(t)$$
$$e_2(t) = E_o \sin \beta a(t) + n_2(t) \qquad (3)$$

In writing these expressions we have assumed that the carrier phase is known and we have assumed that the distortion due to a limited RF bandwidth B is negligible. (This point is considered in more detail in Appendix A.)

2) Convert the continuous time variable to a discrete set of time samples by letting $t = t_o + K\Delta$, $K = 0, 1, \cdots N$ where $\Delta = \frac{1}{B}$. The MMSE estimate is now given by the conditional expectation of the modulation $\langle a \mid e_R \rangle$ It can be determined by evaluating the derivative of the conditional characteristic function $\gamma(i\omega)$ at the origin, $i\omega = 0$.

3) Set up an integral representing the characteristic function $\gamma(i,\omega)$ of the conditional probability density of the modulating signal having the value $a(t)$ given the received signal $e_R(t)$

$$\gamma(i\omega) = K \int_A da \; e^{i\omega a_m + \left(\frac{E_s}{G_n}\right)^2 \left(\frac{n}{E_s}\right) \cdot \cos\left(\beta a - \theta\right) - \frac{1}{2} a \cdot R a^{-1}} \qquad (15)$$

---

*Equation numbers in this summary correspond to the numbers of equations developed later in this Chapter.

where the modulating signal $a(\tau)$ is assumed to be a Gaussian random vector with zero mean and covariance matrix R and a_m is the value of $a(\tau)$, $\tau = m\Delta$. (The covariance matrix R is related to the autocorrelation function $R_a(t)$ by the relation $R_{mn} = R_a((m-n)\Delta)$.) The evaluation of the integral (Equation (15)) is the crucial step in this analysis. If the integral could be analytically evaluated the problem of determining the MMSE estimate could be considered to be completely solved from a theoretical point of view.

4) Approximately evaluate this integral for $\gamma(i\omega)$

   a) By a power series method for low signal-to-noise ratios.

   b) By an asymptotic method for large signal-to-noise ratios.

5) Determine the desired MMSE estimate by evaluating the logarithmic derivative of the characteristic function at the origin and determine the associated conditional minimum mean-square error by evaluating the second derivative of the logarithm of the characteristic function. (Use of the logarithmic derivatives eliminates the need to evaluate the normalizing constant in Equation (15).)

6) Because of the nonlinear nature of the modulation the conditional minimum mean-square error, determined in step 5 is a function of the received signal $e_R(t)$. The final step is to determine the minimum mean-square error by averaging the conditional minimum mean-square error over $e_R(t)$. (The conditional minimum mean-square error would be independent of the received signal with linear modulation when both the noise and the modulating signal are Gaussian processes.)

The results of this analysis are:

For small signal-to-noise ratios $\left(\frac{E_0}{\sigma_n}\right)^2 << 1$, where σ_n^2 is the RF noise power in bandwidth B, we have

$$\hat{a}_m = \beta\left(\frac{E_0}{\sigma_n}\right)^2 e^{-\frac{1}{2}\beta^2\sigma_a^2} \sum_{k=1}^{N} R_{mk}\left(\frac{n_k}{E_0}\right) \sin\theta_k + O\left(\frac{E_0}{\sigma_n}\right)^4 \qquad (21)$$

and

$$\mathcal{E}^2 = \sigma_a^2 - \beta^2\left(\frac{E_0}{\sigma_n}\right)^2 e^{-\frac{1}{2}\beta^2\sigma_a^2} \sum_{k=1}^{N} R_{mk}^2\left(\frac{n_k}{E_0}\right) \cos\theta_k + O\left(\frac{E_0}{\sigma_n}\right)^4 \qquad (22)$$

In the case of bandlimited white modulation and an infinite observation interval we have for the expected value of the error

$$\frac{MSE}{\sigma_a^2} = 1 - \beta^2\sigma_a^2(S/N)_i\, e^{-\beta^2\sigma_a^2} - \frac{\pi\sqrt{3K}}{8}\,\frac{(S/N)_i^2}{\frac{1}{2}\beta^2\sigma_a^2}\left(1 - 2\sqrt{2}\; e^{-\beta^2\sigma_a^2}\right) \tag{29}$$

where $(S/N)_i = \frac{E_o^2}{2N_oW}$

For large signal-to-noise ratios:

$$\hat{a}_m = \tilde{a}_m - \tfrac{1}{2}\,\frac{\partial}{\partial i\omega}\,\ell n\,\|P\| + C(0) \tag{61}$$

$$MSE = P^{-1}_{mm} - \tfrac{1}{2}\,\frac{\partial^2}{\partial i\omega^2}\,\ell n\,\|P\|\Big|_{i\omega=0} + \frac{\partial C}{\partial i\omega}\Big|_{i\omega=0} \tag{62}$$

where $\tilde{a}$ is the solution of the equation

$$R^{-1}X + \frac{E_o^2\beta}{\sigma_n^2}\left[\frac{\pi}{E_o}\sin(\beta X - \theta)\right] = 0 \tag{63}$$

and P is the matrix

$P = R^{-1} + \left(\frac{E_o}{\sigma_n}\right)^2\beta^2\left[\frac{\pi}{E_o}\cos(\beta X - \theta)\right]$, where $\left[\frac{\pi}{E_o}\cos(\beta X - \theta)\right]$ is a diagonal matrix.

We conjecture that the terms $C(0)$ and $\frac{\partial C}{\partial i\omega}$ are negligible for large signal-to-noise ratios and for moderate signal-to-noise ratios when the observation time is long. This has not been rigorously demonstrated and requires further study.

We may note that the first term $\tilde{a}_m$ may be shown to be the equivalent to the output of an unrealizable phase-locked loop (i.e., a PLL which includes nonrealizable filters).

For the case of a flat signal spectrum we obtain the following approximation to the expected value of the mean-square error

$$\frac{\sigma_a^2}{\beta^2 MSE} = \frac{(S/N)_i\,\langle Y\rangle}{1 + \left(\frac{1}{2\langle Y\rangle} + \frac{1}{\langle Y\rangle^2}\right)/(S/N)_i}\,, \quad \beta >> 1$$

where

$$\langle Y\rangle = 1 - \tfrac{3}{2}\beta^2 MSE$$

This result is plotted in Figure II-1.

As discussed in Chapter I, it is known that this result becomes incompatible with bounds derived by Shannon[1] for $\beta \to \infty$. It is believed that this conflict is due to the approximations which were made in evaluating the expected value of the mean-square error.

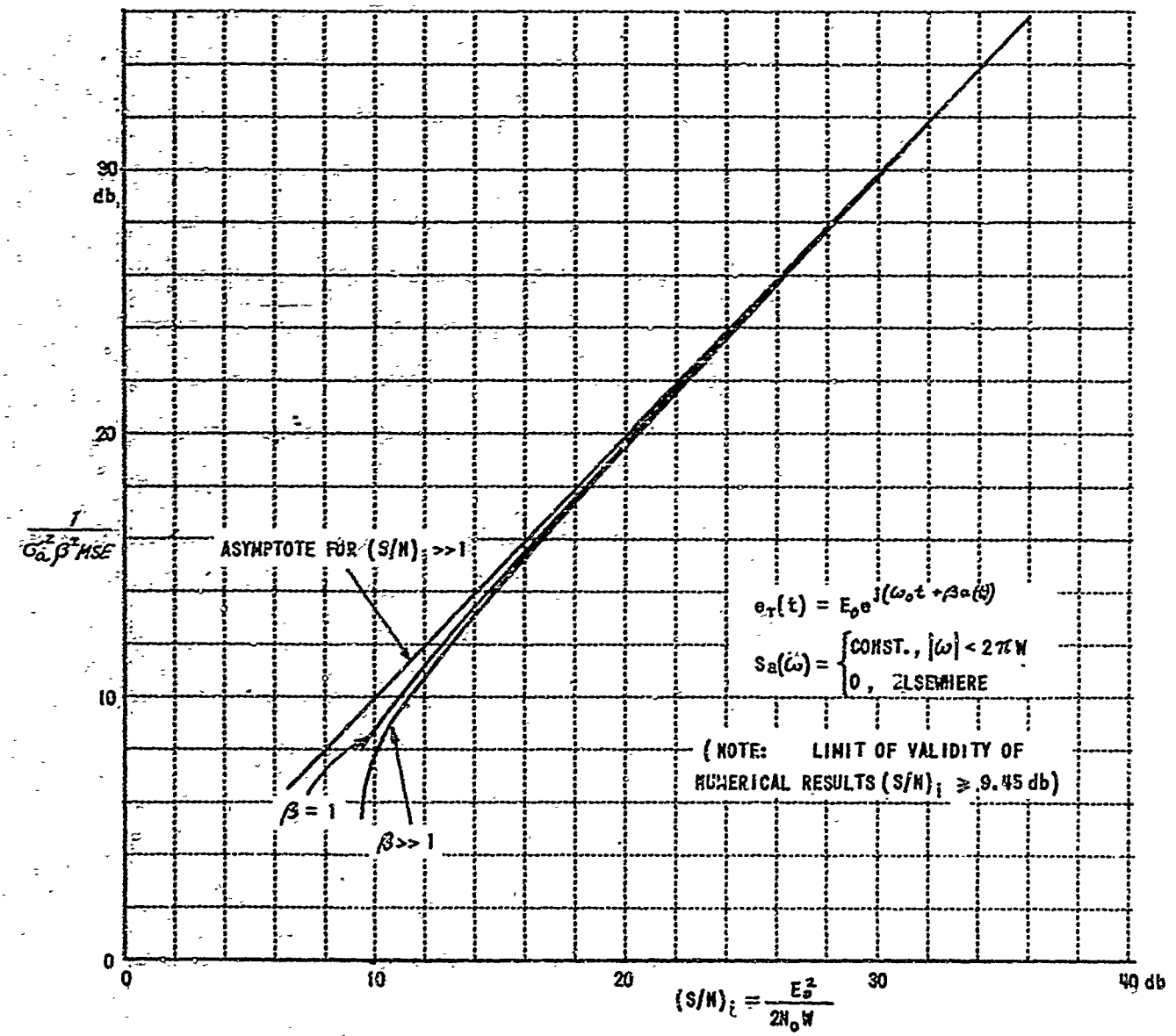


Figure II-1 RESULTS OF ANALYTIC EVALUATION OF MMSE PERFORMANCE OF PM DEMODULATION

2. Development of Minimum Mean-Square Error Estimate

Consider the received waveforms

$$e_R(t) = e_T(t) + n(t) = E_o \cos\left(\omega_o t + \beta a(t)\right) + n(t) \tag{1}$$

where $n(t)$ is an additive noise
E_o is carrier amplitude
ω_o is carrier frequency
$a(t)$ is the modulating signal
β is the index of modulation
t is the time
$e_R(t)$ is the received signal.

Expressing $e_R(t)$ in terms of its in phase and quadrature components:

$$e_R(t) = e_1(t) \cos \omega_o t - e_2(t) \sin \omega_o t \tag{2}$$

where

$$\begin{aligned} e_1(t) &= E_o \cos \beta a(t) + n_1(t) \\ e_2(t) &= E_o \sin \beta a(t) + n_2(t) \end{aligned} \tag{3}$$

Instead of dealing with the continuous input, take a set of equally spaced samples at times $t_h = t_o + h\Delta$, $h = 1, 2, \cdots, N$

Let

$$\begin{aligned} e_1(t_o + h\Delta) &= e_{1k} = E_o \cos \beta a_k + n_{1k} \\ e_2(t_o + h\Delta) &= e_{2k} = E_o \sin \beta a_k + n_{2k} \end{aligned} \tag{4}$$

Thus we reduce the problem to a finite dimensional vector problem

Let e_1 be the vector with components e_{1k}
e_2 be the vector with components e_{2k}
a be the vector with components a_k
n_1 be the vector with components n_{1k}
n_2 be the vector with components n_{2k} (5)

Given the joint probability of the additive noise and the transmitted signal

$$dP = p(a, n_1, n_2)\, da\, dn_1\, dn_2 \quad ; \left\{ \begin{array}{l} \text{since } a, n_1, n_2 \text{ vectors} \\ da = da_1\, da_2 \cdots da_M, \text{ etc.} \end{array} \right. \tag{6}$$

we may make a change of variable from n to e yielding

$$dP = p_0(a,\ e_1 - E_0 \cos\beta a,\ e_2 - E_0 \sin\beta a)\, da\, de_1\, de_2 \tag{7}$$

Now let us make the following assumptions:

The signal a and the noise components n_1, n_2 are independent;

The noise is bandlimited white Gaussian with zero mean and mean-square value $\sigma_n^2 = BN_0$

The noise spectrum is symmetrical about ω_0 and the time interval Δ is selected such that $B\Delta = 1$, where B is the RF bandwidth in Hz.

Then

$$p(a, n_1, n_2) = p(a)\, p(n_1)\, p(n_2)$$

and

$$p(n_1) = \frac{1}{(2\pi\sigma_n^2)^{N/2}}\ e^{-\frac{1}{2\sigma_n^2} n_1 \cdot n_1}$$

$$p(n_2) = \frac{1}{(2\pi\sigma_n^2)^{N/2}}\ e^{-\frac{1}{2\sigma_n^2} n_2 \cdot n_2} \tag{8}$$

where

$$n_1 \cdot n_1 \equiv \sum_{k=1}^{N} n_{1k}^2$$

Thus

$$dP = p(a)\, \frac{1}{(2\pi\sigma_n^2)^N}\ e^{-\frac{1}{2\sigma_n^2}\left[(e_1 - E_0\cos\beta a)\cdot(e_1 - E_0\cos\beta a) + (e_2 - E_0\sin\beta a)\cdot(e_2 - E_0\sin\beta a)\right]} da\, de_1\, de_2 \tag{9}$$

If we integrate equation (9) over a we get the differential probability of the received signal (e_1, e_2). The conditional probability of a signal vector given the received signal would be

$$P(a \mid e_1, e_2)\, da = \frac{p(a, e_1, e_2)\, da}{P(e_1, e_2)} = \frac{dP}{\int_A dP} \tag{10}$$

We know that the conditional expectation of $a(\tau)$ a component of $\mathbf{a}$ will be the minimum mean-square error estimate of the modulation at time τ The problem is to find this value and to compute the mean-square error. The difficulty in the computation is caused by the sinusoidal nonlinearity in equation (9). We are going to attempt to approximate this result by considering an approximation to the conditional characteristic function of a_τ.

Let $dP_a = dP / de_1\, de_2$

The characteristic function $\gamma(i\omega)$ is

$$\gamma(i\omega) = \int dP_a\, e^{i\omega a(\tau)} \tag{11}$$

(τ corresponds to one of the sampled times, i.e., $\tau = t_0 + h\Delta$ for some h in the range considered.)

Now consider the function $\mu(\omega) = \ln \gamma(i\omega)$ and note that

$\hat{a}(\tau) = \frac{\partial}{\partial i\omega}\mu(\omega) = \frac{1}{\gamma(0)} \frac{\partial \gamma(i\omega)}{\partial i\omega}\Big|_{i\omega=0}$ condition expectation of $a(\tau)$ given e_1, e_2.

$$\mathcal{E}^2 = \frac{\partial^2 \mu(\omega)}{\partial i\omega}\Big|_{i\omega=0} = \qquad e_1, e_2 \tag{12}$$

Assume the modulation $\mathbf{a}$ is Gaussian with zero mean and covariance matrix R_a

Thus $$p(\mathbf{a}) = \frac{1}{(2\pi)^{n/2} \|R_a\|^{1/2}}\, e^{-\frac{1}{2}\mathbf{a}\cdot R_a^{-1}\mathbf{a}} \tag{13}$$

where $\|R_a\|$ is the determinant of the matrix R_a

For convenience we make a change to polar coordinates

$$\begin{aligned} e_{1k} &= r_k \cos\theta_k \\ e_{2k} &= r_k \sin\theta_k \\ de_1 de_2 &= r\, dr\, d\theta \end{aligned} \tag{14}$$

Combining equations (9), (11), (13), and (14), we get:

$$\gamma(i\omega) = \frac{1}{(2\pi)^{n/2}\sigma_n^{2n}\|R_a\|^{1/2}} \int_A d\mathbf{a}\, e^{i\omega a(\tau) - \frac{1}{2\sigma_n^2}\left[NE_s^2 + r\cdot r - 2Er\cdot\cos(\beta a - \theta)\right] - \frac{1}{2}\mathbf{a}\cdot R_a^{-1}\mathbf{a}} \tag{15}$$

$$\gamma(i\omega) = K\int_A da\, e^{i\omega a(\tau) + \frac{E_0^2}{\sigma_n^2}\frac{n_\tau}{E_0}\cdot\cos(\beta a-\theta) - \frac{1}{2}a\cdot R_a^{-1}a} \qquad (15)$$
con't.

where K is a constant with respect to a which may be ignored since we only need the logarithmic derivatives of $\gamma(i\omega)$.

So far only the approximations that have been made are the conversion of the continuous data to a discrete set and we have ignored the distortion effects due to bandlimiting. (See Appendix II-A for a discussion of the effects of bandlimiting.)

Small Signal-to-Noise Ratios

As $\left(\frac{E_0}{\sigma_n}\right)^2 \longrightarrow 0$ we may expand $\exp\left(\frac{E_0}{\sigma_n^2}n\cdot\cos(\beta a-\theta)^2\right)$ in a power series and get

$$\gamma(i\omega) = K\int da\left[1+\frac{E_0}{\sigma_n^2}n\cdot\cos(\beta a-\theta)+\frac{1}{2!}\left(\frac{E_0}{\sigma_n^2}n\cdot\cos(\beta a-\theta)\right)^2\cdots\right]e^{i\omega a(\tau)-\frac{1}{2}a\cdot R_a^{-1}a}$$

This integral is evaluated by use of the identity[2]

$$\int\frac{1}{(2\pi^{1/2})\|R_a\|^{1/2}}e^{-\frac{1}{2}a\cdot R_a^{-1}a+i\omega\cdot a}da = e^{-\frac{1}{2}\omega\cdot R_a\omega},$$ yielding:

$$\gamma(i\omega) = K_1\left\{e^{-\frac{1}{2}\omega^2\sigma_a^2}+\left(\frac{E_0}{\sigma_n}\right)^2\sum_{k=1}^{N}\frac{n_k}{E_0}\,\frac{1}{2}\left[e^{-\frac{1}{2}(\omega_\tau+\beta_k)\cdot R_a(\omega_\tau+\beta_k)}+e^{-\frac{1}{2}(\omega_\tau-\beta_k)\cdot R_a(\omega_\tau-\beta_k)+i\theta_k}\right]\right.$$

$$+\frac{1}{2}\left(\frac{E_0}{\sigma_n}\right)^4\sum_{j=1}^{N}\sum_{k=1}^{N}n_k n_j\,\frac{1}{4}\left[e^{-i(\theta_j+\theta_k)-\frac{1}{2}(\omega_\tau+\beta_j+\beta_k)\cdot R_a(\omega_\tau+\beta_j+\beta_k)}\right.$$

$$+e^{i(-\theta_j+\theta_k)-\frac{1}{2}(\omega_\tau+\beta_j-\beta_k)\cdot R_a(\omega_\tau+\beta_j+\beta_k)}$$

$$+e^{-i(-\theta_j+\theta_k)-\frac{1}{2}(\omega_\tau+\beta_j+\beta_k)\cdot R_a(\omega_\tau-\beta_j+\beta_k)}$$

$$\left.\left.+e^{i(\theta_j+\theta_k)-\frac{1}{2}(\omega_\tau-\beta_j-\beta_k)\cdot R_a(\omega_\tau-\beta_j-\beta_k)}\right] \qquad (16)\right.$$

$$+\ O\left(\frac{E_0}{\sigma_n}\right)^6\cdots$$

Where the notation ω_τ is used for a vector with zero components in all positions except for the τ^{TH} where it has the value ω. Similarly, for β_k and σ_a^{-2} is the element on the diagonal of R_a (mean-square value of signal a).

The derivative of Equation (16) is

$$\frac{\partial \gamma}{\partial i\omega} = K_1 \left\{ i\omega \sigma_a^2 \; e^{-\frac{1}{2}\omega^2\sigma_a^2} \right.$$

$$+ \left(\frac{E_0}{\sigma_n}\right)^2 \sum_k \frac{n_k}{E_0} \; \frac{1}{2} \left[\begin{array}{l} u_\tau \cdot R_a(i\omega_\tau + i\beta_k) \; e^{-\frac{1}{2}(\omega_\tau+\beta_k)\cdot R_a(\omega_\tau+\beta_k) + i\theta_k} \\ + u_\tau \cdot R_a(i\omega_\tau + i\beta_k) \; e^{-\frac{1}{2}(\omega_\tau+\beta_k)\cdot R_a(\omega_\tau+\beta_k) + i\theta_k} \end{array} \right]$$

$$+ \tfrac{1}{2}\left(\frac{E_0}{\sigma_n}\right)^4 \sum_k \sum_j \frac{n_k n_j}{E_0} \tfrac{1}{4} \left\{ \left(u_\tau - R_a(i\omega_\tau + \beta_j i + i\beta_k)\right) e^{-i(\theta_j+\theta_k) - \frac{1}{2}(\omega_\tau+\beta_j+\beta_k)\cdot R_a(\omega_\tau+\beta_j+\beta_k)} \right.$$

$$+ \left(u_\tau \cdot R_a(i\omega_\tau + i\beta_j - i\beta_k)\right) e^{i(-\theta_j+\theta_k) - \frac{1}{2}(\omega_\tau+\beta_i-\beta_k)\cdot R_a(\omega_\tau+\beta_j-\beta_k)}$$

$$+ \left(u_\tau \cdot R_a(i\omega_\tau - i\beta_j + i\beta_k)\right) e^{-i(-\theta_j+\theta_k) - \frac{1}{2}(\omega_\tau-\beta_j+\beta_k)\cdot R_a(\omega_\tau-\beta_j+\beta_k)}$$

$$\left. + \left(u_\tau \cdot R_a(i\omega_\tau - i\beta_j - i\beta_k)\right) e^{i(\theta_j+\theta_k) - \frac{1}{2}(\omega_\tau-\beta_j-\beta_k)\cdot R_a(\omega_\tau-\beta_j-\beta_k)} \right\}$$

$$+ O\left(\frac{E_0}{\sigma_n}\right)^6 \qquad (17)$$

where u_τ is used to denote a unit vector with a one in the τ position.

The second derivative evaluated at $i\omega = 0$ is

$$\left.\frac{\partial^2 \gamma}{\partial i\omega^2}\right|_{i\omega=0} = K_1 \left\{ \sigma_a^2 + \left(\frac{E_0}{\sigma_n}\right)^2 \sum_k \frac{n_k \cos\theta_k}{E_0} \left(\sigma_a^2 \, e^{-\frac{1}{2}\beta^2\sigma_a^2} - \beta^2 R_{a\tau k}^2 \, e^{-\frac{1}{2}\beta^2\sigma_a^2}\right) \right.$$

$$+ \tfrac{1}{2}\left(\frac{E_0}{\sigma_n}\right)^4 \sum_k \sum_j \frac{n_k n_j}{E_0^2} \tfrac{1}{4} \left\{ \left(\sigma_a^2 - \left(u_\tau \cdot R_a(\beta_j+\beta_k)\right)^2\right) e^{-i(\theta_j+\theta_k) - \frac{1}{2}(\beta_j+\beta_k)\cdot R_a(\beta_j+\beta_k)} \right.$$

$$+ \left(\sigma_a^2 - \left(u_\tau \cdot R_a(\beta_j-\beta_k)\right)^2\right) e^{i(-\theta_j+\theta_k) - \frac{1}{2}(\beta_j-\beta_k)\cdot R_a(\beta_j-\beta_k)}$$

$$+ \left(\sigma_a^2 - \left(u_\tau \cdot R_a(\beta_j-\beta_k)\right)^2\right) e^{-i(\theta_j+\theta_k) - \frac{1}{2}(\beta_j-\beta_k)\cdot R_a(\beta_j-\beta_k)}$$

$$\left. + \left(\sigma_a^2 - \left(u_\tau \cdot R_a(\beta_j+\beta_k)\right)^2\right) e^{i(\theta_j+\theta_k) - \frac{1}{2}(\beta_j+\beta_k)\cdot R_a(\beta_j+\beta_k)} \right\} \qquad (18)$$

$$+ O\left(\frac{E_0}{\sigma_n}\right)^6$$

Now

$$\gamma(0) = K_1\left\{1 + \left(\frac{E_0}{\sigma_n}\right)^2 \sum_k r_k \cos\theta_k\, e^{-\frac{1}{2}\beta^2\sigma_a^2} + \frac{1}{2}\left(\frac{E_0}{\sigma_n}\right)^4 \sum_k \sum_j \frac{r_k r_j}{E_0^2} \frac{1}{2}\left[\cos(\theta_j+\theta_k)\, e^{-\frac{1}{2}(\beta_j+\beta_k)\cdot R_a(\beta_j+\beta_k)} + \cos(\theta_j+\theta_k) e^{-\frac{1}{2}(\beta_j-\beta_k)\cdot R_a(\beta_j-\beta_k)}\right] + \cdots\right. \tag{19}$$

and

$$\left.\frac{\partial\gamma}{\partial i\omega}\right|_{i\omega=0} = K_1\left\{\left(\frac{E_0}{\sigma_n}\right)^2 \sum r_k \sin\theta_k \beta R_a(K\Delta) e^{-\frac{1}{2}\beta^2\sigma_a^2} + \frac{1}{2}\left(\frac{E_0}{\sigma_n}\right)^4 \sum\sum \frac{r_k r_j}{E_0} \frac{1}{4}\left[iU_\tau \cdot R_a(\beta_j+\beta_k) e^{-i(\theta_j+\theta_k)-\frac{1}{2}(\beta_j+\beta_k)\cdot R_a(\beta_j+\beta_k)} + iU_\tau \cdot R_a(\beta_j-\beta_k) e^{i(-\theta_j+\theta_k)-\frac{1}{2}(\beta_j-\beta_k)\cdot R_a(\beta_j-\beta_k)} - iU_\tau \cdot R_a(\beta_j-\beta_k) e^{-i(-\theta_j+\theta_k)-\frac{1}{2}(\beta_j-\beta_k)\cdot R_a(\beta_j-\beta_k)} - iU_\tau \cdot R_a(\beta_j+\beta_k) e^{i(\theta_j+\theta_k)-\frac{1}{2}(\beta_j+\beta_k)\cdot R_a(\beta_j+\beta_k)}\right]\right. \tag{20}$$

For very small $\left(\frac{E_0}{\sigma_n}\right)^2$ we have

$$\hat{a}(\tau) = \frac{1}{\gamma(0)} \left.\frac{\partial\gamma(\omega)}{\partial i\omega}\right|_{i\omega=0} \approx \beta\left(\frac{E_0}{\sigma_n}\right)^2 e^{-\frac{1}{2}\beta^2\sigma_a^2} \sum_{k=1}^{N} \frac{r_k}{E_0} \sin\theta_k\, R_a(K\Delta) + O\left(\frac{E_0}{\sigma_n}\right)^4 \tag{21}$$

and

$$\varepsilon^2 = \frac{1}{\gamma^2(0)}\left[\gamma(0)\frac{\partial^2\gamma}{\partial i\omega^2} - \left(\frac{\partial\gamma}{\partial i\omega}\right)^2\right]_{i\omega=0} \approx \sigma_a^2 - \beta^2\left(\frac{E_0}{\sigma_n}\right)^2 e^{-\frac{1}{2}\beta^2\sigma_a^2} \sum_{k=1}^{N} \frac{r_k}{E_0} \cos\theta_k R_a^2(K\Delta) + O\left(\frac{E_0}{\sigma_n}\right)^4 \tag{22}$$

Since $r_k \cos\theta_k = E_c \cos\beta a_k + n_{ik}$ and $\langle r_k \cos\theta_k\rangle = e_0 e^{-\frac{1}{2}\beta^2\sigma_a^2}$

The expected value of N_0 may be computed as

$$\text{MSE} = \langle\varepsilon^2\rangle = \sigma_a^2 - \frac{\beta^2 E_0^2}{\sigma_n^2} e^{-\beta^2\sigma_a^2} \sum_{k=1}^{N} R_a(K\Delta) + \cdots \tag{23}$$

Replacing the sum by an integral we get for a final result

$$\text{MSE} = \sigma_a^2 - \frac{\beta^2 E_0^2}{\Delta\sigma_n^2} e^{-\beta^2\sigma_a^2} \int_{t_0}^{t_N} dt\, R_a^2(\tau - t) + \cdots \tag{24}$$

For a flat lowpass signal spectrum of width W Hz and intensity A

$$R_a(\tau)\ R_a(0)\ \frac{\sin 2\pi W\tau}{2\pi W\tau} \tag{25}$$

$$\sigma_a^2 = R_a(0) = WA_o$$

and if $t_o = -\infty$, $t_N = +\infty$

$$\int_{-\infty}^{\infty} d\tau\, R_a^2(\tau - t) = \frac{\sigma_a^2}{2W} = \sigma_a^2 \mathcal{E}_1^2 \tag{26}$$

$$\sigma^2 = BN_o$$

$$\sigma_n^2 = \sigma_a^2$$

$$B\Delta = 1$$

$$(S/N)_i = \frac{E_o^2}{2N_o W}$$

Thus, $$MSE = \sigma_a^2 \left[1 - \beta^2 \frac{E_o^2}{2} \frac{\mathcal{E}_1^2}{N_o} e^{-\beta^2\sigma_a^2} \cdots\cdots \right] \tag{27}$$

where $$\sigma_n^2 = N_o B$$

$$B\Delta = 1$$

substituting

$$= \sigma_a^2 \left[1 - \beta^2 (S/N)_i\, \sigma_a^2\, e^{-\beta^2\sigma_a^2} \cdots\cdots \right]$$

Comparison of Equation (27) with Equation (22) of Chapter III, shows agreement upon use of the relationship between λ_o and σ_a^2/MSE derived in Appendix C of Chapter IV.

Including terms of $\left(E_0/\sigma_n\right)^4$ we have

$$\mathcal{E}^2 = \sigma_a^2 - \frac{E_o}{\sigma_n^2}\beta^2 e^{-\frac{1}{2}\beta^2\sigma_a^2}\sum r_k \cos^2\theta_k R_a^2(K\Delta) \tag{28}$$

$$+\frac{E^2}{4\sigma_n^2}\beta^2 e^{-\beta^2\sigma_a^2}\sum\sum r_j r_k\left\{\cos\theta_j\cos\theta_k\left[4R_a^2(k\Delta)-\left(R_a(K\Delta)+R_{a\tau_j}\right)^2 e^{-\beta^2 R_{ajk}}-\left(R_{a\tau_j}-R_a(K\Delta)\right)^2 e^{\beta^2 R_{ajk}}\right]\right.$$

$$\left.+\sin\theta_j\sin\theta_k\left[-4R_a(K\Delta)R_{a\tau_j}+\left(R_{a\tau_j}+R_a(K\Delta)\right)^2 e^{-\beta^2 R_{ajk}}-\left(R_{a\tau_j}-R_a(K\Delta)\right)^2 e^{\beta^2 R_{ajk}}\right]\right\}.$$

Taking the expected value yields*

$$MSE = \left\langle \mathcal{E}^2 \right\rangle = \sigma_a^2 - \frac{\beta^2 E_o^2}{\sigma_n^2} e^{-\beta^2\sigma_a^2}\sum R_a^2(K\Delta)$$

$$-\beta^2 e^{-2\beta^2\sigma_a^2}\sum_k\sum_j\left[\left(e^{2\beta^2 R_{ajk}}-e^{\beta^2 R_{ajk}}\right)\left(R_{a\tau_j}^2+R_a^2(K\Delta)-2R_{a\tau_j}R_a(K\Delta)\right)\right. \tag{29}$$

$$\left.+\left(e^{-2\beta^2 R_{ajk}}-e^{-\beta^2 R_{ajk}}\right)\left(R_{a\tau_j}^2+R_a^2(K\Delta)+2R_{a\tau_j}R_a(K\Delta)\right)\right]$$

$$+O\left(\frac{E_o}{\sigma_n}\right)^6$$

These may be approximated by integrals.

---

*Note $$\left\langle r_j\cos\theta_j + r_k\cos\theta_k\right\rangle = \frac{E_o^2}{2}e^{-\beta^2\sigma_a^2}\left(e^{\beta^2 R_{ajk}}+e^{-\beta^2 R_{ajk}}\right)+\sigma_n^2\delta_{jk}$$

$$\left\langle r_j\sin\theta_j\; r_j\sin\theta_k\right\rangle = \frac{E_o^2}{2}e^{-\beta^2\sigma_a^2}\left(e^{\beta^2 R_{ajk}}-e^{-\beta^2 R_{ajk}}\right)+\sigma_n^2\delta_{jk}$$

Integral Approximation

$$MSE = \sigma_a^2 - \frac{\beta^2 E_o^2}{\Delta \sigma_n^2} e^{-\beta^2 \sigma_a^2} \int_{t_o}^{t_n} dt\, R_a^2(\tau - t) \tag{28'}$$

$$+ \frac{E_o^4 \beta^2}{4\Delta^2 \sigma_n^4} e^{-2\beta^2 \sigma_a^2} \int_{t_o}^{t_n} dt_1 \int_{t_o}^{t_n} dt_2 \left\{ \left[e^{-\beta^2 R_a(t_1 - t_2)} - e^{-2\beta^2 R_a(t_1 - t_2)} \right] \times \left[R_a(\tau - t_1) + R_a(\tau - t_2) \right]^2 \right.$$

$$\left. + \left[e^{\beta^2 R_a(t_1 - t_2)} - e^{+2\beta^2 R_a(t_1 - t_2)} \right] \times \left[R_a(\tau - t_1) - R_a(\tau - t_2) \right]^2 \right\} + O\left(\frac{E_o}{\sigma_n}\right)^6$$

If we let $R_a = \sigma_a^2 \frac{\sin 2\pi W\tau}{2\pi W\tau}$; $\sigma_a^2 = 2\pi A_o W e^2$

$t_o = -\infty$, $t_N = +\infty$

Assume $\beta^2 \sigma_a^2 > 1$, second term predominates.

Approximately we have

$$\frac{MSE}{\sigma_a^2} \approx 1 - \beta^2 \sigma_a^2 (S/N)_i e^{-\beta^2 \sigma_a^2} - \frac{\pi \sqrt{3\pi}}{8} \frac{(S/N)_i^2}{\sqrt{\beta \sigma_a^2}} \left(1 - 2\sqrt{2}\, e^{-\beta^2 \sigma_a^2}\right) + \cdots$$

where $(S/N)_i = \frac{E_o^2}{2N_o W}$ (29')

For very small $\beta^2 \sigma_a^2 << 1$

$$\frac{MSE}{\sigma_a^2} \approx 1 - \beta^2 \sigma_a^2 (S/N)_i + \frac{4}{\pi} \beta^4 \sigma_a^2 (S/N)_i^2 \tag{30}$$

Large Signal-to-Noise Ratio

Returning to Equation (15)

$$\gamma(i\omega) = k \int_A da\, e^{i\omega a(\tau) + \frac{E_o^2}{\sigma_n^2} \frac{\pi}{E_o} \cdot \cos(\beta a - \theta) - \frac{1}{2} a \cdot R_a^{-1} a} \tag{15}$$

The integral is of the form

$I = \int_A da\, e^{f(a)}$, where a is a vector and $f(a)$ is a scalar function, thus

$$I = \int_{-\infty}^{\infty} da_1 \int_{-\infty}^{\infty} da_2 \cdots \int_{-\infty}^{\infty} da_N \; e^{f(a_1, a_2, \cdots, a_N)} \tag{31}$$

Expand $f(a)$ in a Taylor series

$$f(a) = f(\tilde{a}) + (a-\tilde{a}) \cdot f'(\tilde{a}) + \tfrac{1}{2} (a-\tilde{a}) \cdot f''(\tilde{a})(a-\tilde{a}) + \cdots\cdots\cdots \tag{32}$$

or in multidimensional form

$$f(a_1, a_2, \cdots a_N) = f(a_1, a_2, \cdots a_N) + \sum_k (a_k - \tilde{a}_k) \frac{\partial f}{\partial a_k} + \tfrac{1}{2} \sum\sum (a_l - \tilde{a}_k)(a_l - \tilde{a}_l) \frac{\partial^2 f}{\partial a_k \partial a_l}$$

$$+ \cdots\cdots\cdots$$

Select $\tilde{a}_k$ to be the value of a_k which make*

$$\frac{\partial f}{\partial a_k} = 0 \tag{33}$$

Then for a first approximation

$$I \approx \int da\, e^{f(\tilde{a}) + \frac{1}{2}(a-\tilde{a}) \cdot f''(\tilde{a})(a-\tilde{a})}$$

$$= (2\pi)^{N/2} e^{f(\tilde{a})} \Big/ \sqrt{-\|f''(\tilde{a})\|} \tag{34}$$

where $\| \; \|$ designates the determinant of the matrix.

Correction terms are obtained by using more terms in the Taylor series.

*The technique we are using is a multidimensional form of the principle of Steepest Descent or Stationary Phase which is used to evaluate asymptotic forms.

$$I \approx \int da\; e^{f(\tilde{a}) + \frac{1}{2}(a-\tilde{a})\cdot f''(\tilde{a})(a-\tilde{a}) + \frac{1}{3!}\sum\sum\sum (a_j-\tilde{a}_j)(a_k-\tilde{a}_k)(a_\ell-\tilde{a}_\ell)\frac{\partial^3 f}{\partial a_j \partial a_k \partial a_\ell}}$$

$$+ \frac{1}{4!}\sum\sum\sum\sum (\quad\quad)$$

$$I \approx \int da\; e^{f(\tilde{a}) + \frac{1}{2}(a-\tilde{a})\cdot f''(\tilde{a})(a-\tilde{a})} \left(1 + \frac{1}{3!}\sum\sum\sum (a_j-\tilde{a}_j)(a_k-\tilde{a}_k)(a_\ell-\tilde{a}_\ell)\frac{\partial^3 f}{\partial a_j \partial a_k \partial a_\ell}\right.$$

$$\left. + \frac{1}{4!}\sum\sum\sum\sum (a_j-\tilde{a}_j)(a_k-\tilde{a}_k)(a_\ell-\tilde{a}_\ell)(a_m-\tilde{a}_m)\frac{\partial^4 f}{\partial a_j \partial a_k \partial a_\ell \partial a_m} + \text{-----}\right) \quad (35)$$

$$+ \text{-------}$$

$$\approx \frac{(2\pi)^{N/2} e^{f(\tilde{a})}}{\sqrt{-\|f''(\tilde{a})\|}} + e^{f(\tilde{a})} \int_A da\; e^{\frac{1}{2}(a-\tilde{a})\cdot f''(\tilde{a})(a-\tilde{a})} \left(\frac{\sum\sum\sum\sum}{4!}(a_j-\tilde{a}_j)(a_k-\tilde{a}_k)\right.$$

$$\left.(a_\ell-\tilde{a}_\ell)(a_m-\tilde{a}_m)\frac{\partial f^4}{\partial a_j\, \partial a_k\, \partial a_\ell\, \partial a_m}\right) \quad (36)$$

(The odd moments are zero because of symmetry)

Now

$$f(a) = i\omega a(\tau) + \frac{E_o^2}{\sigma_n^2}\frac{n}{E_o}\cdot\cos(\beta a - \theta) - \tfrac{1}{2}\, a\cdot R_a^{-1} a \quad (37)$$

$$\frac{\partial f(a)}{\partial a_k} = i\omega\delta_{k\pi}\frac{E_o^2\beta}{\sigma_n^2}\frac{n_k}{E_o}\sin(\beta a_k - \theta_k) - u_k\cdot R_a^{-1} a \quad (38)$$

where u_k is a unit vector in the k^{TH} position and $\delta_{k\tau}$ unless k corresponds to τ when $\delta_{k\tau}=1$

$$\frac{\partial^2 f(a)}{\partial a_k \partial a_\ell} = -\frac{E_o^2\beta}{\sigma_n^2}\frac{n_k}{E_o}\cos(\beta a_k - \theta_k)\delta_{k\ell} - u_k\cdot R_a^{-1} u_\ell \quad (39)$$

$$\frac{\partial^3 f}{\partial a_k \partial a_\ell \partial a_m} = \frac{E_o^2\beta^3}{\sigma_n^2}\frac{n_k}{E_o}\sin(\beta a_k - \theta_k)\delta_{k\ell m} \quad = 0, \text{ unless } k=\ell=m \quad (40)$$

$$\frac{\partial^4 f}{\partial a_k \partial a_\ell \partial a_m \partial a_n} = \frac{E_o^2 \beta^4}{\sigma_n^2} \frac{\pi_k}{E_o} \cos(\beta a_k - \theta_k) \delta_{k\ell mn} \quad = 0, \text{ unless } k=\ell=m=n \tag{41}$$

Thus, about the point $\tilde{a}$

$$f(a) = f(\tilde{a}) + \tfrac{1}{2}(a-\tilde{a}) \cdot \left(-R_a^{-1} - \frac{E_o^2 \beta^2}{\sigma_n^2} \left\{ \frac{\pi}{E_o} \cos(\beta \tilde{a} - \theta) \right\} \right) (a-\tilde{a})$$
$$+ \tfrac{1}{3!} \sum_1^N (a_k - \tilde{a}_k)^3 \frac{\partial^3 f}{\partial a_k^3} + \tfrac{1}{4!} \sum_1^N (a_k - \tilde{a}_k)^4 \frac{\partial^4 f}{\partial a_k^4} + \text{-------} \tag{42}$$

where $\left\{ \frac{\pi}{E_o} \cos(\beta \tilde{a} - \theta) \right\}$ is a diagonal matrix with elements $\frac{\pi_k}{E_o} \cos(\beta \tilde{a}_k - \theta_k)$ on the diagonal.

Define the matrix P as

$$P = R_a^{-1} + \frac{E_o^2 \beta^2}{\sigma_n^2} \left\{ \frac{\pi}{E_o} \cos(\beta \tilde{a} - \theta) \right\} \tag{43}$$

Note the point $\tilde{a}$ is defined by the matrix equation

$$R_a^{-1} \tilde{a} + \frac{E_o^2 \beta}{\sigma_n^2} \left[\frac{\pi}{E_o} \sin(\beta \tilde{a} - \theta) \right] = i\omega \delta_{k\tau} \tag{44}$$

where $\left[\frac{\pi}{E_o} \sin(\beta \tilde{a} - \theta) \right]$ is a vector with components $\frac{\pi_k}{E_o} \sin(\beta \tilde{a}_k - \theta_k)$

We have

$$\gamma(i\omega) \approx K \int_A da\, e^{f(\tilde{a}) - \frac{1}{2}(a-\tilde{a}) \cdot P (a-\tilde{a}) + \frac{1}{3!} \sum (a_k - \tilde{a}_k) \frac{\partial^3 f}{\partial a_k^3}\Big|_{a=\tilde{a}} + \frac{1}{4!} \sum (a_k - \tilde{a}_k)^4 \frac{\partial^4 f}{\partial a_k^4}\Big|_{a=\tilde{a}} + \cdots}$$

$$\approx K e^{f(\tilde{a})} \left\{ \frac{(2\pi)^{N/2}}{\sqrt{\|P\|}} + \int_A da\, e^{-\frac{1}{2}(a-\tilde{a}) \cdot P (a-\tilde{a})} \left\{ \tfrac{1}{3!} \sum (a_k - \tilde{a}_k)^3 \frac{\partial^3 f}{\partial a_k^3} + \tfrac{1}{4!} \sum (a_k - \tilde{a}_k)^4 \right. \right.$$
$$\left. \left. \frac{\partial^4 f}{\partial a_k^4} + \text{-----} \right\} \right. \tag{45}$$

where we have used

$$\int_A da\, e^{-\frac{1}{2}(a-\tilde{a}) \cdot P (a-\tilde{a})} = \frac{(2\pi)^{N/2}}{\sqrt{\|P\|}}, \quad \text{(if } P \text{ is positive definite)} \tag{46}$$

Also we have $\int_A da\, e^{-\frac{1}{2}(a-\tilde{a})\cdot P(a-\tilde{a})}(a_k-a_k)^3 = 0$ from symmetry

and $\int_A da\, e^{-\frac{1}{2}(a-\tilde{a})\cdot P(a-\tilde{a})}(a_k-\tilde{a}_k)^4 = \frac{(2\pi)^{N/2}}{\sqrt{\|P\|}}\frac{3}{\lambda_k^2}$; where λ_k (47)

is the k^{TH} eigenvalue of P.

Thus

$$\gamma(i\omega) \approx \frac{K e^{f(\tilde{a})}(2\pi)^{N/2}}{\sqrt{\|P\|}}\left[1+\frac{3}{4!}\sum_{k=1}^{N}\frac{E_o^2\beta^4}{\sigma_n^2}\frac{r_k\cos(\beta\tilde{a}_k-\theta_k)}{E_o\,\lambda_k^2}+\cdots\right] \tag{48}$$

$$\mu(\omega)=\ln\gamma(i\omega)=\ln K(2\pi)^{N/2}+f(\tilde{a})-\tfrac{1}{2}\ln\|P\|+\ln\left(1+\frac{3}{4!}\frac{E_o^2\beta^4}{\sigma_n^2}\sum_1^N\frac{r_k\cos(\beta\tilde{a}_k-\theta_k)}{E_o\,\lambda_k^2}+\cdots\right) \tag{49}$$

$$\frac{\partial\mu(\omega)}{\partial i\omega}=\frac{\gamma\ln(i\omega)}{\partial i\omega}=\frac{\partial f(\tilde{a})}{\partial i\omega}-\frac{1}{2}\frac{\partial\|P\|}{\partial i\omega}+\frac{-\frac{3}{4!}\frac{E_o^2\beta^5}{\sigma_n^2}\sum_1^N\frac{r_k\sin(\beta\tilde{a}_k-\theta_k)}{E_o\,\lambda_k^2}\frac{\partial\tilde{a}_k}{\partial i\omega}+\cdots}{1+\frac{3}{4!}\frac{E_o^2\beta^4}{\sigma_n^2}\sum_1^N\frac{r_k\cos(\beta\tilde{a}_k-\theta_k)}{E_o\,\lambda_k^2}+\cdots} \tag{50}$$

The MMSE estimate is thus given by

$$\hat{a}(\tau)=\left.\frac{\partial\mu(\omega)}{\partial i\omega}\right|_{i\omega=0} \tag{51}$$

Let

$$C(\omega)=\frac{-\frac{3}{4!}\frac{E_o^2\beta^5}{\sigma_n^2}\sum_1^N\frac{r_k\sin(\beta\tilde{a}_k-\theta_k)}{E_o\,\lambda_k^2}\frac{\partial\tilde{a}_k}{\partial i\omega}+\cdots}{1+\frac{3}{4!}\frac{E_o^2\beta^4}{\sigma_n^2}\sum_1^N\frac{r_k\cos(\beta\tilde{a}_k-\theta_k)}{E_o\,\lambda_k^2}+\cdots} \tag{52}$$

Thus

$$\frac{\partial\mu(\omega)}{\partial i\omega}=\frac{\partial f}{\partial i\omega}-\frac{1}{2}\frac{\partial\ln\|P\|}{\partial i\omega}+C(\omega) \tag{53}$$

and

$$\frac{\partial^2\mu(\omega)}{\partial i\omega^2}=\frac{\partial^2 f}{\partial i\omega^2}-\frac{1}{2}\left(\frac{\partial}{\partial i\omega}\right)^2\ln\|P\|+\frac{\partial C(\omega)}{\partial i\omega} \tag{54}$$

Since at $a = \hat{a}$

$$f(\tilde{a}) = i\omega \tilde{a}(T) + \left(\frac{E_o}{\sigma_n}\right)^2 \frac{r}{E_o} \cdot \cos\left(\beta\tilde{a} - \theta\right) - \tfrac{1}{2}\,\tilde{a} \cdot R_a^{-1}\tilde{a} \tag{55}$$

$$\frac{\partial f(\tilde{a})}{\partial i\omega} = \tilde{a}(T) + i\omega \frac{\partial \tilde{a}(T)}{\partial i\omega} - \frac{E_o^2\beta}{\sigma_n^2} \sum \frac{r_k}{E_o} \sin\left(\beta\tilde{a}_k - \theta_k\right) \frac{\partial \tilde{a}_k}{\partial i\omega} - \tilde{a} \cdot R_a^{-1} \frac{\partial \tilde{a}}{\partial i\omega} \tag{56}$$

and since

$$R_a^{-1}\tilde{a} + \frac{E_o^2\beta}{\sigma_n^2}\left[\frac{r}{E_o} \sin\left(\beta\tilde{a} - \theta\right)\right] = i\omega\,\delta_{k\tau} \tag{57}$$

$$R_a^{-1} \frac{\partial \tilde{a}}{\partial i\omega} + \frac{E_o^2\beta^2}{\sigma_n^2}\left\{\frac{r}{E_o} \cos\left(\beta\tilde{a} - \theta\right)\right\} \frac{\partial \tilde{a}}{\partial i\omega} = \delta_{k\tau} \tag{58}$$

$$\frac{\partial f(\tilde{a})}{\partial i\omega} = \tilde{a}(\tau) \tag{59}$$

$$\frac{\partial^2 f(\tilde{a})}{\partial i\omega^2} = \frac{\partial \tilde{a}(\tau)}{\partial i\omega} \tag{60}$$

The MMSE estimate is then given by

$$\hat{a} = \frac{\partial \mu}{\partial i\omega}\bigg|_{i\omega=0} = \tilde{a}(\tau) - \tfrac{1}{2} \frac{\partial}{\partial i\omega} \ln \|P\|\bigg|_{i\omega=0} + C(0) \tag{61}$$

and the mean-square error by

$$\text{MSE} = \frac{\partial^2 \mu}{\partial i\omega^2}\bigg|_{i\omega=0} = \frac{\partial \tilde{a}(\tau)}{\partial i\omega}\bigg|_{i\omega=0} - \tfrac{1}{2} \frac{\partial^2}{\partial i\omega^2} \ln \|P\|\bigg|_{i\omega=0} + \frac{\partial C}{\partial i\omega}\bigg|_{i\omega=0} \tag{62}$$

where $\tilde{a}(\tau)$ is the solution of

$$R_a^{-1} a + \frac{E_o^2\beta}{\sigma_n^2}\left[\frac{r}{E_o} \sin\left(\beta a - \theta\right)\right] = 0 \quad \text{for the element corresponding to the } \tau \text{ time position} \tag{63}$$

and since $\frac{\partial \tilde{a}}{\partial i\omega} = P^{-1}\delta_{k\tau}$ (64)

$$\frac{\partial \tilde{a}(\tau)}{\partial i\omega} = \left(P^{-1}\right)_{\tau\tau}, \quad \text{the diagonal element corresponding to the } \tau \text{ time position} \tag{65}$$

Consideration of Correction Term

Consider $C(\omega)$ as defined by equation (52)

The denominator is

$$\mathcal{D} = 1 + \frac{3}{4!} \frac{E_0^2 \beta^4}{\sigma_n^2} \sum_1^N \frac{n_k}{E_0} \frac{\cos(\beta \tilde{a}_k - \theta_k)}{\lambda_k^2} + \cdots . \tag{66}$$

but (see Appendix C)

$$\begin{aligned} n_k \cos(\beta \tilde{a}_k - \theta_k) &= n_k \cos\beta(\tilde{a}_k - a_k) \cos(\beta a_k - \theta_k) - \lambda_k \sin\beta(\tilde{a}_k - a_k) \sin(\beta a_k - \theta_k) \\ &= (E_0 + n'_{1k}) \cos\beta(\tilde{a}_k - a_k) - n'_{2k} \sin\beta(\tilde{a}_k - a_k) \end{aligned} \tag{67}$$

If we assume $\left(\beta(\tilde{a}_k - a_k)\right) << 1$, we have approximately

$$n_k \cos\beta(\tilde{a}_k - \theta_k) \approx E_0 + n'_{1k} \tag{68}$$

with an expected value of E_0

Therefore

$$\langle \mathcal{D} \rangle \approx 1 + \frac{3}{4!} \frac{E_0^2 \beta^4}{\sigma_n^2} \sum_1^N \frac{1}{\lambda_k^2} + \cdots . \tag{69}$$

Consider the matrix P

$$P = R_a^{-1} + \frac{E_0^2 \beta^2}{\sigma_n^2} \left\{ \frac{n}{E_0} \cos(\beta \tilde{a} - \theta) \right\} \tag{70}$$

With a similar approximation to the one above

$$\langle P \rangle = R_a^{-1} + \frac{E_0^2 \beta^2}{\sigma_n^2} I \tag{71}$$

$$\langle P \rangle^{-1} \approx \left(R_a^{-1} + \frac{E_0^2 \beta^2}{\sigma_n^2} I \right)^{-1} \tag{72}$$

Approximately

$$\langle P \rangle^{-1}_{mn} = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, e^{i\omega\Delta(m-n)} \frac{\varphi(\omega)}{1 + \frac{E_0^2 \beta^2}{\sigma_n^2} \varphi(\omega)} \tag{73}$$

where $\varphi(\omega)$ is the Fourier transform of the correlation R_a

and

$$\langle P \rangle_{mn}^{-2} \approx \frac{1}{\Delta} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \; e^{i\omega\Delta(m-n)} \left(\frac{\varphi(\omega)}{1 + \frac{E_0^2/\beta^2}{\Delta\sigma_n^2} \varphi(\omega)} \right)^2 \tag{74}$$

The $\sum_1^N \frac{1}{\lambda_n^2}$ will be the trace of $\langle P \rangle^{-2}$.

hence

$$\begin{aligned} \sum_1^N \frac{1}{\lambda_n^2} &= \sum_{m=1}^{N} \frac{1}{\Delta} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \left(\frac{\varphi(\omega)}{1 + \frac{E_0^2/\beta^2}{\Delta\sigma_n^2} \varphi(\omega)} \right)^2 \\ &= \frac{1}{\Delta^2} N\Delta \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \left(\frac{\varphi(\omega)}{1 + \frac{E_0^2/\beta^2}{\Delta\sigma_n^2} \varphi(\omega)} \right)^2 \end{aligned} \tag{75}$$

The term $N\Delta$ represents the observation interval. Thus as

$N\Delta \to \infty$, $\langle \mathcal{P} \rangle \to \infty$

The numerator is

$$\begin{aligned} \eta &= -\frac{3}{4} \frac{E_0^2/\beta^5}{\sigma_n^2} \sum \frac{n_k \sin(\beta\tilde{a}_k - \theta_k)}{E_0 \lambda_k^2} \frac{\partial \tilde{a}_k}{\partial i\omega} + \cdots \\ &\approx -\frac{3}{4} \frac{E_0^2/\beta^5}{\sigma_n^2} \sum \frac{n_k \sin\beta(\tilde{a}_k - a_k)\cos(\beta a_k - \theta_k) + n_k \cos\beta(\tilde{a}_k - a_k)\sin(\beta a_k - \theta_k)}{E_0 \lambda_k^2} \frac{\partial \tilde{a}_k}{\partial i\omega} \end{aligned}$$

$$\eta \approx -\frac{3}{4} \frac{E_0^2/\beta^5}{\sigma_n^2} \sum \frac{\sin\beta(\tilde{a}_k - a_k)(E_0 + n'_{1k}) + \cos\beta(\tilde{a}_k - a_k) n'_{2k}}{E_0 \lambda_k^2} \frac{\partial \tilde{a}_k}{\partial i\omega} \tag{76}$$

Since $\frac{\partial \tilde{a}_k}{\partial i\omega}$ behaves as the mean-square error

MSE and if $|\beta(\tilde{a}_k - a_k)| << 1$

$$\eta \approx -\frac{3}{4} \frac{E_0^2/\beta^5}{\sigma_n^2} \sum \frac{\left\langle \beta(\tilde{a}_k - a_k) n'_{1k} \frac{\partial \tilde{a}_k}{\partial i\omega} \right\rangle + \left\langle n'_{2k} \frac{\partial \tilde{a}_k}{\partial i\omega} \right\rangle}{E_0 \lambda_k^2} \tag{77}$$

$$\langle \eta \rangle \approx -\frac{3}{4} \frac{E_0^2/\beta^5}{\sigma_n^2} \sum \frac{\left\langle \beta(\tilde{a}_k - a_k) n'_{1k} \frac{\partial \tilde{a}_k}{\partial i\omega} \right\rangle + \left\langle n'_{2k} \frac{\partial a_k}{\partial i\omega} \right\rangle}{E_0 \lambda_k^2} \tag{78}$$

we conjecture that this term is small compared to the denominator.

$$\left\langle \frac{n}{D} \right\rangle \to 0 \qquad N\Delta \to \infty .$$

This argument is not intended to be rigorous but only serves to suggest the plausibility of neglecting $c(\omega)$.

Evaluation of Logarithmic Derivatives

Consider

$$\left\| P \right\| = \left\| R_{\tilde{a}}^{-1} + \frac{E_0 \beta^2}{\sigma_n^2} \left\{ \frac{n}{E_0} \cos\left\{ (\beta\tilde{a} - \theta) \right\} \right\} \right\| \tag{79}$$

$$\frac{1}{\|P\|} \frac{\partial \|P\|}{\partial i\omega} = \sum_k \frac{\partial \|P\|}{\|P\| \partial \tilde{a}_k} \frac{\partial \tilde{a}_k}{\partial i\omega} \tag{80}$$

but we have

$$\frac{\partial \tilde{a}_k}{\partial i\omega} = P_{k\tau}^{-1}$$

and

$$\frac{1}{\|P\|} \frac{\partial \|P\|}{\partial a_k} = -\frac{E_0 \beta^3}{\sigma_n^2} \frac{n_k}{E_0} \sin\left(\beta\tilde{a}_k - \theta_k\right) \frac{P_{kk}}{\|P\|} = -\frac{E_0}{\sigma_n^2} \frac{n_k}{E_0} \sin\left(\beta\tilde{a}_k - \theta_k\right) P_{kk}^{-1}$$

where P_{kk} is a principal minor of P

Thus the correction to the MMSE estimator is

$$-\frac{1}{2} \frac{\partial \ln \|P\|}{\partial i\omega} = \frac{E_0^2 \beta^3}{2\sigma_n^2} \sum_k P_{kk}^{-1} P_{k\tau}^{-1} \frac{n_k}{E_0} \sin\left(\beta\tilde{a}_k - \theta_k\right) \tag{81}$$

For the contribution to the error we get

$$-\frac{1}{2} \frac{\partial^2 \ln \|P\|}{\partial i\omega^2} = \frac{E_0 \beta^3}{2\sigma_n^2} \sum_k \left\{ \beta \frac{n_k}{E_0} \cos\left(\beta\tilde{a}_k - \theta_k\right) P_{k\tau}^{-1} P_{kk}^{-1} P_{k\tau}^{-1} + \frac{n_k}{E_0} \sin\left(\beta\tilde{a}_k - \theta_k\right) \left[P_{n\tau}^{-1} \frac{\partial P_{kk}^{-1}}{\partial i\omega} - \left| P_{kk}^{-1} \frac{\partial P_{k\tau}^{-1}}{\partial i\omega} \right| \right] \right\} \tag{82}$$

To summarize we have shown that the MMSE estimate

$$\hat{a} = \tilde{a}_\tau - \frac{1}{2} \frac{\partial}{\partial i\omega} \ln \|P\| + c(0) \tag{83}$$

and the conditional MMSE

$$\mathcal{E}^2 = P_{\tau\tau}^{-1} - \frac{1}{2} \frac{\partial^2}{\partial i\omega^2} \ln \|P\| + \left. \frac{\partial c}{\partial i\omega} \right|_{i\omega = 0} \tag{84}$$

where $\tilde{a}_\tau$ is the solution of

$$R_a^{-1}\tilde{a} + \frac{E_o^2\beta}{\sigma_n^2}\left[\frac{n}{E_o}\sin(\beta\tilde{a}-\theta)\right] = 0 \tag{85}$$

$$P = R_a^{-1} + \frac{E_o^2\beta^2}{\sigma_n^2}\left\{\frac{n}{E_o}\cos(\beta\tilde{a}-\theta)\right\} \tag{86}$$

$$-\frac{1}{2}\frac{\partial \ln\|P\|}{\partial i\omega} = \frac{E_o^2\beta^3}{2\sigma_n^2}\sum_k P_{kk}^{-1}P_{k\tau}^{-1}\frac{n_k}{E_o}\sin(\beta\tilde{a}_k-\theta_k) \tag{87}$$

$$\begin{aligned}-\frac{1}{2}\frac{\partial^2 \ln\|P\|}{\partial i\omega^2} &= \frac{E_o\beta^3}{2\sigma_n^3}\sum_k \beta\frac{n_k}{E_o}\cos(\beta\tilde{a}_k-\theta_k)\left(P_{k\tau}^{-1}\right)^2 P_{kk}^{-1} \\ &+ \frac{E_o^2\beta^3}{2\sigma_n^2}\sum_k \frac{n_k}{E_o}\sin(\beta\tilde{a}_k-\theta_k)\left[P_{k\tau}^{-1}\frac{\partial P_{kk}^{-1}}{\partial i\omega} + P_{kk}^{-1}\frac{\partial P_{k\tau}^{-1}}{\partial i\omega}\right]\end{aligned} \tag{88}$$

$C(0)$ and $\frac{\partial C}{\partial i\omega}$ are assumed to be negligible (this point requires further investigation).

Note that $\mathcal{E}^2$ is a function of the particular received signal since P is a function of $\tilde{a}$. Hence we shall attempt to evaluate the expected value of $\mathcal{E}^2$ to obtain a measure of the performance of an MMSE estimator.

In evaluating the integral in Equation (46) it was necessary to assume that the matrix $[P]$ is positive definite. Since $[P]$ is a stochastic matrix the validity of the results for the average of $\mathcal{E}^2$ to be obtained, requires that probability of $[P]$ not being positive definite, contributes negligible error to the evaluation of the MMSE.

Approximation to Obtain Numerical Results

Consider the matrix $P = R_a^{-1} + \frac{E_o\beta^2}{\sigma_n^2}\left\{n\cos(\beta\tilde{a}-\theta)\right\}$ (89)

$$\begin{aligned}&= R_a^{-1} + \frac{E_o\beta^2}{\sigma_n^2}\left\langle n\cos(\beta\tilde{a}-\theta)\right\rangle + \frac{E_o\beta^2}{\sigma_n^2}\left\{n\cos(\beta\tilde{a}-\theta) - \left\langle n\cos(\beta\tilde{a}-\theta)\right\rangle\right\} \\ &\equiv T^{-1} + Q\end{aligned}$$

where

$$T^{-1} = R_a^{-1} + \frac{E_0\beta^2}{\sigma_n^2}\left\{\left\langle n\cos(\beta\tilde{a}-\theta)\right\rangle\right\} \tag{90}$$

$$Q = \frac{E_0\beta^2}{\sigma_n^2}\left\{n\cos(\beta\tilde{a}-\theta) - \left\langle n\cos(\beta\tilde{a}-\theta)\right\rangle\right\} \tag{91}$$

Under certain restrictions* the inverse matrix P^{-1} may be obtained from the expansion

$$P^{-1} = T - TQT + TQTQT - \cdots\cdots \tag{92}$$

Thus (since $\langle Q\rangle = 0$)

$$\left\langle P^{-1}\right\rangle = T + T\left\langle QTQ\right\rangle T + \cdots \tag{93}$$

Let us make the further simplifying assumption that the diagonal matrix $\left\{\left\langle n\cos(\beta\tilde{a}-\theta)\right\rangle\right\}$ is a constant times the identity matrix I.

Thus

$$T^{-1} = \sigma_n^{-1} + \frac{E_0^2\beta^2}{\sigma_n^2}\langle X\rangle I \tag{94}$$

where $\langle X\rangle = \langle X_m\rangle = \left\langle \frac{n_m}{E_0}\cos(\beta\tilde{a}_m-\theta_m)\right\rangle$ (95)

For very large signal-to-noise ratios $\langle X\rangle \sim 1$

An element of T may now be approximated by

$$T_{mn} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{i\omega\Delta(m-n)}\,\frac{\varphi(\omega)}{1+\frac{E_0^2\beta^2}{\Delta\sigma_n^2}\langle X\rangle\varphi(\omega)} \tag{96}$$

Where $\varphi(\omega)$ is the power density spectrum of the modulating signal.

Thus, the dominant term in the error is

$$\left\langle P^{-1}\right\rangle_{\tau\tau} \sim T_{\tau\tau} + \cdots\cdots \simeq \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\,\frac{\varphi(\omega)}{1+\frac{E_0^2\beta^2}{\Delta\sigma_n^2}\langle X\rangle\varphi(\omega)} \tag{97}$$

*Expansion will be valid when the norm of $T^{1/2}QT^{1/2}$ is less than one.

For a constant $\varphi(\omega)$,

$$\varphi(\omega) = \frac{A_0}{2} \quad ; \quad |\omega| < 2\pi W \tag{98}$$

$$\varphi(\omega) = 0 \quad , \quad |\omega| > 2\pi W \tag{99}$$

$$\frac{1}{2\pi} \int d\omega \, \frac{\varphi(\omega)}{1 + \frac{E_0^2 \beta^2}{\Delta \sigma_n^2} \langle x \rangle \varphi(\omega)} = \frac{\sigma_a^2}{1 + Z}$$

where $\sigma_a^2 = WA_0$ the mean-square value of the modulating signal

and $$Z = \frac{E_0^2 \beta^2 A_0}{2N_0} \langle x \rangle = \beta^2 \sigma_a^2 (S/N)_i \langle x \rangle \tag{100}$$

where $$(S/N)_i = \frac{E_0^2}{2\sigma_n^2} \times \frac{B}{W} = \frac{E_0^2}{2N_0 W} \tag{101}$$

Consider the term

$$J = \langle T Q T Q T \rangle_{\tau\tau}$$

$$J = \sum_{n_1, n_2, n_3, n_4} \langle T_{\tau n_1} Q_{n_1 n_2} T_{n_2 n_3} Q_{n_3 n_4} T_{n_4 \tau} \rangle \tag{102}$$

Since Q is diagonal

$$J = \sum_{m,n} T_{\tau m} \langle Q_{mm} Q_{nn} T_{mn} T_{n\tau} \rangle \tag{103}$$

Let $$\langle Q_{mm} Q_{nn} \rangle = \left(\frac{E_0^2 \beta^2}{\sigma_n^2} \right)^2 \mu_{mn} \tag{104}$$

where $$\mu_{mn} = \langle (x_m - \bar{x}_m)(x_n - \bar{x}_n) \rangle \tag{105}$$

Approximating by integrals*

$$J \approx \sum_{m,n} \left(\frac{E_0^2 \beta^2}{\sigma_n^2} \right)^2 \frac{1}{(2\pi)^3} \int du_1 \int du_2 \int du_3 \, e^{iu_1 \Delta(\tau - m) + iu_2 \Delta(m - n) + iu_3 \Delta(n - \tau)} f(u_1) \, f(u_2) \, f(u_3) \tag{106}$$

---

*Appendix B describes the relationship between integrals and the corresponding sums.

where $f(u) = \varphi(u) \Big/ \left(1 + \frac{E_o^2 \beta^2}{\Delta \sigma_n^2} \langle x \rangle \varphi(u)\right)$ (107)

If we further assume

$$\mu_{mn} = 0 \qquad m \neq n$$

$$\mu_{mn} = \mu_2 \qquad m = n$$

$$J \sim \left(\frac{E_o^2 \beta^2}{\sigma_n^2}\right)^2 \sum_m \frac{1}{(2\pi)^3} \int du_1 \int du_2 \int du_3 \; e^{i u_1 \Delta(\tau - m) + i u_3 \Delta(n - \tau)} f(u_1)\, f(u_2)\, f(u_3)\, \mu_2$$

$$\sim \left(\frac{E_o^2 \beta^2}{\sigma_n^2}\right)^2 \mu_2 \frac{1}{\Delta} \left(\frac{1}{2\pi} \int du_1 f^2(u_1)\right) \left(\frac{1}{2\pi} \int du_2 f(u_2)\right) \tag{108}$$

For constant $\varphi(\omega)$

$$J \sim \left(\frac{E_o^2 \beta^2}{\sigma_n^2}\right)^2 \frac{\mu_2}{\Delta} \frac{\sigma_a^2}{1+Z} \frac{E_o^2 \sigma_a^2}{(1+Z)^2}$$

$$\sim \frac{\sigma_a^2}{1+Z} \left[\frac{\beta^2 \sigma_a^2 Z}{(1+Z)^2} \frac{\mu_2}{\langle x \rangle} \frac{E_o^2}{\sigma_n^2}\right] \tag{109}$$

Thus the evaluation of $\langle P^{-1} \rangle_{\tau\tau}$ to the second term is

$$\langle P^{-1} \rangle_{\tau\tau} \sim \frac{\sigma_a^2}{1+Z} \left\{1 + \frac{\beta^2 \sigma_a^2 Z}{(1+Z)^2} \frac{\mu_2}{\langle x \rangle} \frac{E_o^2}{\sigma_n^2} + \cdots\right\} \tag{110}$$

For very large $(S/N)_i$ ratio

$$\langle x \rangle \sim 1 \quad , \mu_2 \frac{E_o^2}{\sigma_n^2} \sim 1$$

Consider the first term in the approximation

$$J_1 \sim -\frac{1}{2} \frac{\langle \partial^2 \ell_n \| n \| \rangle}{\partial i \omega^2} \sim \frac{E_o \beta^3}{2\sigma_n^2} \sum_k \left\langle \beta r_k \cos(\beta \tilde{a}_k - \theta_k) (P_{k\tau}^{-1})^2 P_{kk}^{-1} \right\rangle + \cdots \tag{111}$$

Approximately replace P^{-1} by T, $r_k \cos(\beta \tilde{a}_k - \theta_k)$ by $x E_o$ and assume we may use the product of the expected values

Then

$$J_1 \sim \frac{E_o^2 \beta^4}{2\sigma_n^2} \langle x \rangle \sum_k (T_{k\tau})^2 (P_{kk}^{-1})$$

$$\sim \frac{E_o^2 \beta^4}{2\sigma_n^2 \Delta} \langle x \rangle \left(\frac{1}{2\pi} \int du_1 f^2(u_1)\right) \langle P^{-1} \rangle_{\tau\tau} \tag{112}$$

Collecting results

$$\left\langle \varepsilon^2 \right\rangle \approx \left\langle P^{-1} \right\rangle_{\tau\tau} + \left\langle P^{-1} \right\rangle_{\tau\tau} \frac{E_0^2 \beta^4}{2\sigma_n^2 \Delta} \langle X \rangle \frac{1}{2\pi} \int du_1 f^2(u_1)$$

$$\approx \left\langle P_{\tau\tau}^{-1} \right\rangle \left(1 + \frac{E_0^2 \beta^4}{2\sigma_n^2 \Delta} \langle X \rangle \frac{1}{2\pi} \int du_1 f^2(u_1) \right) \tag{113}$$

For a constant $\varphi(\omega)$

$$\left\langle \varepsilon^2 \right\rangle \approx \left\langle P_{\tau\tau}^{-1} \right\rangle \left(1 + \frac{\beta^2 \sigma_a^2 Z}{2(1+Z)^2} \right) \tag{114}$$

where

$$\left\langle P^{-1} \right\rangle_{\tau\tau} \approx \frac{\sigma_a^2}{1+Z} \left\{ 1 + \frac{\beta^2 \sigma_a^2 Z}{(1+Z)^2} \frac{\mu_2}{\langle X \rangle} \frac{E_0^2}{\sigma_n^2} + \dots \right\} \tag{110}$$

$$Z \equiv E_0^2 \beta^2 \frac{A_0}{2N_0} \langle X \rangle = \beta^2 \sigma_a^2 (S/N)_i \langle X \rangle \tag{100}$$

$$(S/N)_i \approx \frac{E_0^2}{2\sigma_n^2} \frac{B}{W} = \frac{E_0^2}{2N_0 W} \tag{101}$$

Consider now the term

where $\langle X \rangle = \left\langle \frac{r}{E_0} \cos(\beta\tilde{a} - \theta) \right\rangle$ (95)

But

$$\frac{r}{E_0} \cos(\beta\tilde{a} - \theta) = \frac{r}{E_0} \cos\left(\beta(\tilde{a} - a) + \beta a - \theta\right)$$

$$= \frac{r}{E_0} \cos\beta(\tilde{a} - a) \cos(\beta a - \theta) - \frac{r}{E_0} \sin\beta(\tilde{a} - a) \sin(\beta a - \theta)$$

$$= \cos\beta(\tilde{a} - a) \left(1 + \frac{n_1}{E_0}\right) - \left(\sin\beta(\tilde{a} - a)\right) \frac{n_2}{E_0}$$

viz Appendix C,

where $\tilde{a} - a$ is the error.

To a first approximation assume

$\tilde{a} - a$ is Gaussian, zero mean, variance $\left\langle \varepsilon^2 \right\rangle$

n_1 is independent of $\tilde{a} - a$

$\left\langle (\tilde{a} - a) n_2 \right\rangle \approx E_0 \beta \left\langle \varepsilon^2 \right\rangle$

we then have $\langle X \rangle \approx e^{-\frac{\beta^2\langle\varepsilon^2\rangle}{2}} - \beta^2\langle\varepsilon^2\rangle \approx 1 - \frac{3}{2}\beta^2\langle\varepsilon^2\rangle$ (115)

These approximations will be valid in the range $\beta^2\langle\varepsilon^2\rangle \ll 1$

To obtain numerical results let us further assume that $\mu_2 = \frac{\sigma_n^2}{E_o^2}$

We then have for the mean-square error

$$\langle\varepsilon^2\rangle = \frac{\sigma_a^2}{1+Z}\left(1 + \frac{\beta^2\sigma_a^2 Z}{(1+Z)^2}\left(\frac{1}{2} + \frac{1}{\langle x\rangle}\right) + \cdots\right) \quad (116)$$

where

$$Z = \beta^2\sigma_a^2(S/N)_i\langle x\rangle \quad (100)$$

$$(S/N)_i = \frac{E_o^2}{2\sigma_n^2}\frac{B}{W} = \frac{E_o^2}{2N_o W} \quad (101)$$

Thus

$$\beta^2\langle\varepsilon^2\rangle = \frac{\beta^2\sigma_a^2}{1+Z}\left(1 + \frac{\beta^2\sigma_a^2}{1+Z}\frac{Z}{1+Z}\left(\frac{1}{2} + \frac{1}{\langle x\rangle}\right) + \cdots\right) \quad (116')$$

In the range where $Z \gg 1$ replace $1+Z$ with Z

$$\beta^2\langle\varepsilon^2\rangle \approx \frac{1}{(S/N)_i\langle x\rangle}\left(1 + \frac{1}{(S/N)_i\langle x\rangle}\left(\frac{1}{2} + \frac{1}{\langle x\rangle}\right)\right) \quad (117)$$

(We may use this expression and (115) to compute the relation between $(S/N)_i$ and $\langle x\rangle$).

Note that $\langle x\rangle$ and $\frac{1}{\beta^2\langle\varepsilon^2\rangle}$ are functions of $(S/N)_i$ only and independent of β (to a first approximation). This approximation is sufficient to show the threshold for all values of $\beta \geq 4$.

Figure II-1 was computed in this manner and shows the dependence of $\frac{1}{\sigma_a^2\beta^2 MSE}$ on $(S/N)_i$ for $\beta = 1$, $\beta \gg 1$ and the linear asymptotic valid for large $(S/N)_i$.

Extension to FM

Although all the analytical work has concentrated on the phase modulated case, it is easily extended to the FM case by the following method.

In FM instead of using $\beta a(t)$ as the modulating function, use $\beta\int_{t_0}^{t} d\tau\, a(\tau)$. In the discrete case

$$\beta\int_{t_0}^{t} d\tau\, a(\tau) \sim \Delta\beta \sum_{l=0}^{k} a_l \tag{118}$$

Equation (15) becomes

$$\delta(i\omega) = K\int_{A} d\omega\; e^{i\omega a_k + \frac{E_0^2}{\sigma_n^2}\sum_k \frac{n_k}{E_0}\cos\left(\beta\Delta\sum_0^k a_k - \theta_k\right) - \frac{1}{2}a\cdot R_a^{-1}a} \tag{119}$$

Let $q_k = \Delta\sum_0^k a_l$

$a_k = (q_k - q_{k-1})/\Delta \quad , \; k = 1, 2, \cdots, N$

$a_0 = q_0/\Delta$

In vector form

$$a = Qq$$

Thus

$$\delta(i\omega) = K_1\int_{q} dq\; e^{i\frac{\omega}{\Delta}(q_\tau - q_{\tau-1}) + \frac{E_0}{\sigma_n^2}n\cdot\cos(\beta q_k - \theta) - \frac{1}{2}q\cdot Q' R_a^{-1} Q q} \tag{120}$$

As in the PM case, for large signal-to-noise ratios let the exponent be represented by the function f.

Then

$$\frac{\partial f}{\partial q_{1k}} = \frac{i\omega}{\Delta}\left[\delta_{k\tau} - \delta_{k\tau-1}\right] - \frac{E_0\beta}{\sigma_n^2}n_k \sin(\beta q_k - \theta_k) - u_k\cdot Q' R_a^{-1} Q q = 0 \qquad k=0,1,\cdots,N \tag{38'}$$

(correspond to equation (38)

$$\frac{\partial^2 f}{\partial q_{1k}\partial q_{1\ell}} = -\frac{E_0\beta^2}{\sigma_n^2}n_k\cos(\beta q_k - \theta_k)\delta_{k\ell} - u_k\cdot Q' R_a^{-1} Q u_\ell \tag{39'}$$

(correspond to equation (39)

$$\frac{\partial F}{\partial i\omega} = \frac{q_\tau - q_{\tau-1}}{\Delta} + \sum_k \frac{\partial F}{\partial q_k}\frac{\partial q_k}{\partial i\omega} \qquad (121)$$

$$= \frac{q_\tau - q_{\tau-1}}{\Delta}$$

when q is selected as the solution of (38') with $i\omega = 0$.

Hence the MMSE estimator corresponding to $\hat{a}_\tau$ of the PM case is

$$\frac{q_{1\tau} - q_{1\tau-1}}{\Delta}$$

Where $q_{1\tau}$ is the solution of the equation

$$Q' R_2^{-1} Q q + \frac{E_0^2\beta^2}{\sigma_n^2}\left[\frac{r}{E_0} - \sin(\beta q - \theta)\right] = 0 \qquad (63')$$

The matrix P of PM is replaced by

$$P = Q' R_a^{-1} Q + \frac{E_0^2\beta^2}{\sigma_n^2}\left\{\frac{r}{E_0} - \sin(\beta q - \theta)\right\} \qquad (43')$$

For the first term in the error $\mathcal{E}^2$ we get

$$\frac{\partial^2 f}{\partial i\omega^2} = \frac{1}{\Delta}\left[\frac{\partial q_\tau}{\partial i\omega} - \frac{\partial q_{\tau-1}}{\partial i\omega}\right] \qquad (122)$$

By differentiating 38' we have

$$P\frac{\partial q}{\partial i\omega} = \frac{u_\tau - u_{\tau-1}}{\Delta}$$ where u_γ is a unit vector in the τ position

Hence

$$\frac{\partial^2 f}{\partial i\omega^2} = \frac{1}{\Delta^2}\left(u_\tau - u_{\tau-1}\right)\cdot P^{-1}\left(u_\tau - u_{\tau-1}\right) \tag{123}$$

which corresponds to $\partial^2 f / \partial i\omega^2 = P^{-1}_{\tau\tau}$ as the first term of $\mathcal{E}^2$

In the case of very large signal-to-noise ratio when we may replace $\langle n \cos(\beta q - \theta)\rangle$ by E_0, we had in PM

$$\langle P\rangle = R_a^{-1} + \frac{E_0^2\beta^2}{\sigma_n^2} I \tag{124}$$

going to an integral approximation we had

$$\langle \xi^2\rangle \approx \left\langle P^{-1}_{\tau\tau}\right\rangle \approx \frac{1}{2\pi}\int d\omega \frac{\varphi(\omega)}{1+\frac{E_0^2\beta^2}{\Delta\sigma_n^2}\varphi(\omega)} \text{ for PM} \tag{125}$$

Now in FM we note that Q is an approximate differentiation as is the vector $\frac{u_\tau - u_{\tau-1}}{\Delta}$ hence for FM

$$\langle \xi^2\rangle \approx \frac{1}{2\pi}\int d\omega \frac{\varphi(\omega)}{1+\frac{E_0^2\beta^2}{\Delta\sigma_n^2}\frac{\varphi(\omega)}{\omega^2}} \qquad \text{for FM}$$

These expressions for $\frac{\partial^2 f}{\partial i\omega^2}$ represent the expected value of $\langle \mathcal{E}^2\rangle$ or MMSE for large signal-to-noise ratios.

APPENDIX II-A

THE EFFECT OF LIMITED RF BANDWIDTH

Consider a filter with a frequency response $F(\omega)$ where:

$$F(\omega) = 1 \;, \quad |\omega \pm \omega_0| \leq 2\pi \frac{B}{2}$$
$$= 0 \;, \quad \text{elsewhere} \tag{A-1}$$

Let the input be $g(t)$, then the output $h(t)$ will be

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, F(\omega) e^{i\omega t} \int_{-\infty}^{\infty} d\tau \, e^{-i\omega\tau} g(\tau)$$
$$h(t) = \int_{-\infty}^{\infty} d\tau \, g(\tau) \, 2 \cos \omega_0 (t-\tau) \, \frac{\sin \pi B (t-\tau)}{\pi (t-\tau)} \tag{A-2}$$

Special Case:

If $\int_{-\infty}^{\infty} d\tau \, e^{-i\omega\tau} g(\tau) = 0$ for all ω where $F(\omega) = 0$ then the filter has no effect, i.e., $h(t) = g(t)$

This will also be true when the spectrum of g is 0 where $F(\omega) = 0$.

Consider a signal $g(t)$ of the form

$$g(t) = E_0 \cos\left(\omega_0 t + L(a(t))\right) + n(t) \tag{A-3}$$

where $n(t)$ is bandlimited Gaussian noise,

$a(t)$ is a random signal,

ω_0 is the carrier radian frequency

E_0 is the carrier amplitude

$L(a)$ is a linear operation on $a(t)$

If the signal $g(t)$ passes through a linear filter with response as given by (A-1) and if the spectrum of the noise is constrained in the bandpass region of the filter, the output $h(t)$ will be:

$$h(t) = n(t) + \int_{-\infty}^{\infty} d\tau \, E_0 \cos\left(\omega_0 \tau + L(a(t))\right) 2 \cos \omega_0 (t-\tau) \, \frac{\sin \pi B (t-\tau)}{\pi (t-\tau)} \tag{A-4}$$

Letting $n(t) = n_1(t)\cos\omega_0 t - n_2(t)\sin\omega_0 t$

and $h(t) = h_1(t)\cos\omega_0 t - h_2(t)\sin\omega_0 t$, We Have (A-5)

$$h_1(t) = n_1(t) + \int_{-\infty}^{\infty} d\tau\, E_0 \left[\cos L(a(t)) + \cos\left(2\omega_0\tau + L(a(t))\right)\right] \frac{\sin\pi B(t-\tau)}{\pi(t-\tau)}$$

$$h_2(t) = n_2(t) + \int_{-\infty}^{\infty} d\tau\, E_0 \left[\sin L(a(t)) - \sin\left(2\omega_c\tau + L(a(t))\right)\right] \frac{\sin\pi B(t-\tau)}{\pi(t-\tau)} \quad \text{(A-6)}$$

If the carrier frequency, ω_0, is very large compared with $2\pi B$ contribution from $\begin{matrix}\cos\\ \sin\end{matrix}\left\{2\omega_0\tau + L(a(t))\right\}$ may be neglected when studying the low frequency output of the filter. Hence we may use

$$h_1(t) = n_1(t) + \int_{-\infty}^{\infty} d\tau\, E_0 \cos L(a(t)) \frac{\sin\pi B(t-\tau)}{\pi(t-\tau)}$$

$$h_2(t) = n_2(t) + \int_{-\infty}^{\infty} d\tau\, E_0 \sin L(a(t)) \frac{\sin\pi B(t-\tau)}{\pi(t-\tau)} \quad \text{(A-7)}$$

If the spectrum of $\begin{matrix}\cos\\ \sin\end{matrix}\left\{L(a(t))\right\}$ is mostly contained in the bandpass of the filter (i.e., if $B/F_L \gg 1$ (where F_L is the effective upper limit of the spectrum of $\begin{matrix}\cos\\ \sin\end{matrix}\left\{L(a(t))\right\}$) then we have

$$h_1(t) = n_1(t) + E_0 \cos\left(L(a(t))\right)$$

$$h_2(t) = n_2(t) + E_0 \sin\left(L(a(t))\right) \quad \text{(A-8)}$$

The difference between equation (A-8) and equation (A-7) represents distortion due to the bandpass filter.

APPENDIX II-B

RELATIONSHIP OF SUMS TO INTEGRALS

Consider a matrix T where

$$T = (I + \alpha R_a)^{-1} \tag{B-1}$$

and where R_a is a covariance matrix

Let $R_a(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{i\omega\tau}\varphi(\omega)$; $\varphi(\omega)$ is a power density spectrum. (B-2)

Then the elements of the covariance matrix will be

$$R_{mn} = R_a((m-n)\Delta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{i\omega\Delta(m-n)}\varphi(\omega) \tag{B-3}$$

The matrix equation $(I + \alpha R_a)X = Y$ (B-4)

may be written $X_m + \alpha \sum_{n=1}^{N} R_{mn} X_n = y_m \quad , m = 1, 2, \ldots, N$ (B-5)

Replacing the sum by an integral

$$X_m + \frac{\alpha}{\Delta}\sum R_{mn} X_n \Delta = y_m$$

$$X(t) + \frac{\alpha}{\Delta}\int d\tau\, x(\tau) R_a(t-\tau) = y(t) \quad \text{where} \quad X(m\Delta) = X_m \tag{B-6}$$

Let $\hat{X}$ be the Fourier transform of X and $\hat{y}$ for y.

Then $\hat{x} + \frac{\alpha}{\Delta}\hat{x}\,\varphi(\omega) = \hat{y}$

solving for $\hat{x}$ (B-7)

$$\hat{x} = \frac{\hat{y}}{1 + \frac{\alpha}{\Delta}\varphi(\omega)}$$

Inverting gives

$$x(t) = \frac{1}{2\pi}\int d\omega e^{i\omega t}\,\frac{\hat{y}}{1+\frac{\alpha}{\Delta}\varphi(\omega)}$$

$$= \frac{1}{2\pi}\int d\omega \left(e^{i\omega t}\Big/ 1+\frac{\alpha}{\Delta}\varphi(\omega)\right)\int dt e^{-i\omega\tau} y(\tau)$$

$$x_i(t) = \int d\tau\, y(t)\,\frac{1}{2\pi}\int d\omega\;\frac{e^{i\omega(t-\tau)}}{1+\frac{\alpha}{\Delta}\varphi(\omega)} \qquad \text{(B-8)}$$

This corresponds to the vector solution

$$x = (I+R_\alpha)^{-1} y = T y$$

$$x_m = \sum T_{mn}\, y_n \qquad \text{(B-9)}$$

From (B-8) we have

$$x(m\Delta) = x_m \cong \sum \Delta\, y_n\,\frac{1}{2\pi}\int d\omega\;\frac{e^{i\omega\Delta(m-n)}}{1+\frac{\alpha}{\Delta}\varphi(\omega)} \qquad \text{(B-10)}$$

Hence the approximation T_{mn} is

$$T_{mn} = \frac{\Delta}{2\pi}\int d\omega\;\frac{e^{i\omega\Delta(m-n)}}{1+\frac{\alpha}{\Delta}\varphi(\omega)} \qquad \text{(B-11)}$$

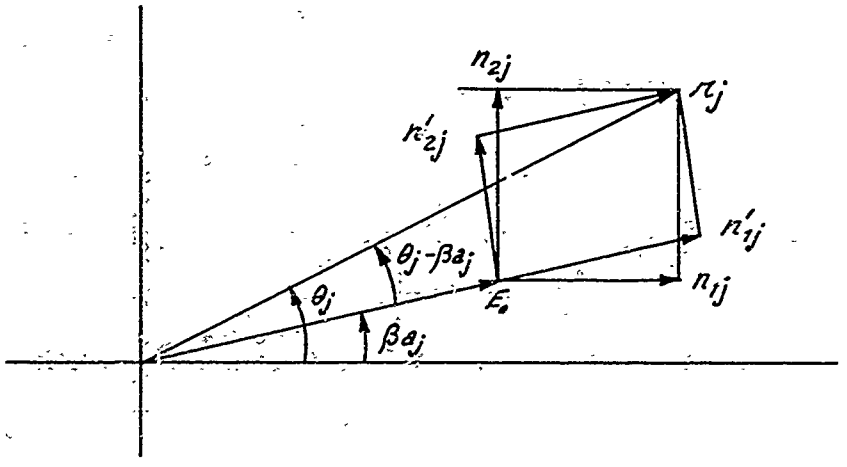
$$= \frac{\Delta}{2\pi}\int d\omega e^{i\omega\Delta(m-n)}\,T(\omega)$$

APPENDIX II-C

ROTATION OF COORDINATES

Consider the equations at time t_j for the in-phase and quadrature comonents.

$$r_j \cos\theta_j = E_o \cos\beta a_j + n_{1j}$$
$$r_j \sin\theta_j = E_o \sin\beta a_j + n_{2j} \qquad \text{(C-1)}$$



Consider the quantities

$$r_j \cos(\theta_j - \beta a_j) = E_o + n_{1j} \cos\beta a_j + n_{2j} \sin\beta a_j$$
$$r_j \sin(\theta_j - \beta a_j) = -n_{1j} \sin\beta a_j + n_{2j} \cos\beta a_j \qquad \text{(C-2)}$$

Let

$$n'_{1j} = n_{1j} \cos\beta a_j + n_{2j} \sin\beta a_j$$
$$n'_{2j} = n_{1j} \sin\beta a_j + n_{2j} \cos\beta a_j \qquad \text{(C-3)}$$

This is an orthoganal transformation. If the statistics of a are independent of the statistics of n the statistics of n' will be the same as the statistics of n .

REFERENCES

[1] Shannon, C. E. and Weaver, W., "The Mathematical Theory of Communication", University of Illinois Press, Urbana, 1949.

[2] Rice, S. O. "Mathematical Analysis of Random Noise", contained in "Noise and Stochastic Processes", Edited by Nelson Wax, Dover Publications, 1954.

III. MINIMUM MEAN SQUARE ERROR ESTIMATION OF RANDOM ANGLE MODULATED SIGNALS; WITH EMPHASIS ON THE LOW $(S/N)_i$ REGION

1. SUMMARY

Section 2 presents a modification of the development in Reference [1] of the MMSE estimate of random PM and FM in the presence of white Gaussian additive noise. The result is expressed as a quotient of two infinite series of operations on the received signal. MMSE estimates are given for both the case of known RF reference phase Φ as well as uniformly randomly distributed Φ. (RF frequency, transmitted amplitude and statistical characteristics of modulation and noise are all assumed to be precisely known at the receiver.)

In Section 3 these MMSE estimates are approximated by the leading term in the series (hence valid for sufficiently small input (RF) signal-to-noise ratios). Expressions for the output signal-to-noise ratios in this asymptotic regime are then derived. It is found that if a very long observation interval is used it is sufficient to restrict one's attention to the case of known Φ in order to characterize the asymptotic output S/N.

Final results show output S/N for PM and FM (both zero delay and infinite delay cases) as a function of input signal-to-noise ratio, modulation characteristic, and modulation index. Typical graphs are included and comparison is made with large signal-to-noise asymptotic results.

2. MINIMUM MEAN SQUARE ERROR (MMSE) ESTIMATE

Random Phase Modulation

Let the modulation a(t) be a zero-mean unit-variance stationary Gaussian process with covariance function R_a and let n(t) be bandlimited white Gaussian noise with spectral density N_o watts/Hz (one-sided spectrum) limited to a band of width 2Ω centered about ω_o (the carrier radian frequency). Let Φ be a uniformly distributed random phase;

furthermore consider a(t), n(t), and Φ to be mutually independent. The received signal is then assumed to be:

$$e_R(t) = E_0 \cos [\omega_0 t + \Phi + \beta a(t)] + n(t) \tag{1}$$

where E_0, ω_0, and β are known positive constants (the last being the modulation index). Assume, furthermore, that Ω is chosen so that $[\omega_0 - \Omega, \omega_0 + \Omega]$ encompasses essentially the full signal power spectrum as well. One can associate 2 Ω with receiver IF bandwidth.

The problem is to find the MMSE estimate of $a(\tau)$, the modulation at time τ, based on knowledge of the received signal over a finite time interval, $\{e_R(t); 0 \le t \le T\}$ or e_R for short. Denote such an estimate by $\hat{a}(e_R, \tau)$ (or $\hat{a}(\tau)$ where e_R is understood).

We assume throughout that $T >> \frac{1}{2\Omega}$ and $\Omega << \omega_0$ (narrow band signals). It is then shown that (Ref. [1]):

$$\hat{a}(\tau) = \langle a(\tau) | e_R \rangle = \frac{\langle a(\tau) Q(e_R, \Phi, a) \rangle_{\Phi, a}}{\langle Q(e_R, \Phi, a) \rangle_{\Phi, a}} \tag{2}$$

where

$$Q(e_R, \Phi, a) \cong \exp\left\{ -\frac{1}{N_0} \int_0^T [e_R(s) - E_0 \cos(\omega_0 s + \Phi + \beta a(s))]^2 ds \right\}$$

and where the subscripts Φ, a indicate that expectations are taken only over the joint distribution of Φ and $a = \{a(s); 0 \le s \le T\}$ with e_R regarded as a given function.

By application of the Karhunen-Loéve expansion for a

$$a(\tau) = \sum_{i=1}^{\infty} \lambda_i^{1/2} A_i \varphi_i(\tau) \; ; \; 0 \le \tau \le T \tag{3}$$

where $\{\lambda_n\}$, $\{\varphi_n\}$ are the (positive) eigenvalues and corresponding eigenfunctions of the integral operator with kernel $\hat{R}_a$ (on $[0,T] \times [0,T]$) and where $\{A_n\}$ are independent standard normal random variables. We can then write, in place of (2),

$$\hat{a}(\tau) = \sum_{i=1}^{\infty} \lambda_i^{1/2} \hat{A}_i \varphi_i(\tau) \; ; \; 0 \le \tau \le T \tag{4}$$

where

$$\hat{A}_i = \frac{\langle A_i Q \rangle_{\Phi,a}}{\langle Q \rangle_{\Phi,a}} \tag{5}$$

The narrow-band assumption permits certain simplifications to be made in the form of Q, viz.,

$$Q \cong q e^{P} = q e^{-\frac{E_0}{N_0} \int_0^T z(s) \cos(\Phi + \beta a(s) + \theta(s))\, ds} \tag{6}$$

where the factor q is independent of a, Φ and where (z, θ) is a normalized magnitude - phase representation of the baseband information in the received signal, viz.,

$$e_R(t) = y_0(t) \cos \omega_0 t + y_1(t) \sin \omega_0 t$$

$$z(t) = \frac{\sqrt{y_0^2(t) + y_1^2(t)}}{E_0}, \quad \theta(t) = \operatorname{sgn}(y_1(t)) \cdot \cos^{-1}\left[\frac{y_0(t)}{E_0 z(t)}\right]$$

Consequently (5) becomes

$$\hat{A}_i \cong \frac{\langle A_i e^{P} \rangle_{\Phi,a}}{\langle e^{P} \rangle_{\Phi,a}} \tag{7}$$

Next, expanding e^P as a power series in P and interchanging order of expectation and summation we obtain

$$\hat{A}_i \cong \frac{\sum_{j=1}^{\infty} \frac{1}{j!} \langle A_i P^j \rangle_{\Phi,a}}{1 + \sum_{k=1}^{\infty} \frac{1}{k!} \langle P^k \rangle_{\Phi,a}} \tag{8}$$

We address ourselves to the task of evaluating the individual expectations in (8). Note, first, by (3) and by the independence of $\Phi, A_1, A_2, \ldots,$ that

$$\langle \cdot \rangle_{\Phi,a} = \lim_{l\to\infty} \langle\langle \cdots \langle\langle \cdot \rangle_{\Phi} \rangle_{A_1} \cdots \rangle_{A_{l-1}} \rangle_{A_l} \tag{9}$$

Next, the following assertions are easily verified:

$$\left\langle \begin{matrix}\cos\\ \sin\end{matrix}(\alpha+\gamma A_i)\right\rangle_{A_i} = e^{-\frac{1}{2}\gamma^2}\begin{Bmatrix}\cos\alpha\\ \sin\alpha\end{Bmatrix} \quad i=1,2,\ldots \tag{10}$$

$$\left\langle A_i \begin{matrix}\cos\\ \sin\end{matrix}(\alpha+\gamma A_i)\right\rangle_{A_i} = \gamma e^{-\frac{1}{2}\gamma^2}\begin{Bmatrix}-\sin\alpha\\ \cos\alpha\end{Bmatrix} \quad i=1,2,\ldots \tag{11}$$

$$\prod_{j=1}^{n}\cos\alpha_j = 2^{-n}\sum_{\nu_i=\pm1,\ i=1,\ldots,n}\cos\Big(\sum_{l=1}^{n}\nu_l\alpha_l\Big)\ ;\ n=1,2,\ldots \tag{12}$$

*

From the definition of P in (6) and identity in (12) we can show that:

$$P_n = \left(\frac{E_0^2}{2N_0}\right)^n \int_0^T\cdots\int_0^T \left[\prod_{l=1}^{n} z(s_l)\,ds_l\right] \sum_{j=0}^{n}\binom{n}{j} \cos\left[(2j-n)\Phi + \sum_{m=1}^{j}(\beta a(s_m)+\theta(s_m)) - \sum_{m=j+1}^{n}(\beta a(s_m)+\theta(s_m))\right]$$

---

* The notation $\sum_{\nu_i=\pm1,\ i=1,\ldots,n}$ means that a summation is to be taken over all 2^n possible combinations of $\nu_i = \pm 1,\ i=1,\ldots,n$. For example,

$$\prod_{j=1}^{2}\cos\alpha_j = 2^{-2}\sum_{\nu_i=\pm1,\ i=1,2}\cos\Big(\sum_{l=1}^{2}\nu_l\alpha_l\Big) = \tfrac{1}{4}\{\cos(\alpha_1+\alpha_2)+\cos(\alpha_1-\alpha_2)+\cos(-\alpha_1+\alpha_2)+\cos(-\alpha_1-\alpha_2)\}$$

$$= \tfrac{1}{2}\{\cos(\alpha_1+\alpha_2)+\cos(\alpha_1-\alpha_2)\}$$

Consequently, by (9) and the permissible interchange of integration and expectation:

$$\left\langle A_i^k \cos^n \right\rangle_{\Phi_k} = \begin{cases} \left(\frac{E_0^2}{2N_0}\right)^n \int_0^T \cdots \int_0^T \left[\prod_{l=1}^{n} \lambda(s_l)\, ds_l\right] \binom{n}{n/2} \left\langle A_i^k \cos\left(\sum_{m=1}^{n} \mu_{m,n}\left(\beta a(s_m) + \theta(s_m)\right)\right)\right\rangle_a ; & n \text{ even} \\ 0 \; ; & n \text{ odd} \end{cases}$$

$$\mu_{m,n} = \begin{cases} 1 \; ; & 1 \le m \le n/2 \\ -1 \; ; & n/2 < m \le n \end{cases}$$

$$k = 0, 1 \qquad (13)$$

(the Φ expectation being zero whenever $2j - n \neq 0$).

Next, (3), (9), (10) [applied repeatedly], (11), and Mercer's Theorem yield

$$\left\langle \cos\left(\sum_{m=1}^{n} \mu_{m,n}\left(\beta a(s_m) + \theta(s_m)\right)\right)\right\rangle_a$$

$$= \cos\left(\sum_{m=1}^{n} \mu_{m,n}\, \theta(s_m)\right) \exp\left\{-\frac{\beta^2}{2} \sum_{j=1}^{\infty} \lambda_j \left(\sum_{k=1}^{n} \mu_{k,n}\, \varphi_j(s_k)\right)^2\right\}$$

$$= \cos\left(\sum_{m=1}^{n} \mu_{m,n}\, \theta(s_m)\right) \exp\left\{-\frac{\beta^2}{2} \sum_{p,q=1}^{n} \mu_{p,n}\, \mu_{q,n}\, R_a(s_p - s_q)\right\}$$

$$n = 1, 2, \ldots, \qquad (14)$$

Similarly,

$$\left\langle A_i \cos\left(\sum_{m=1}^{n} \mu_{m,n}\left(\beta a(s_m) + \theta(s_m)\right)\right)\right\rangle_a$$

$$= -\beta \lambda_i^{1/2} \left(\sum_{l=1}^{n} \mu_{l,n}\, \varphi_i(s_l)\right) \sin\left(\sum_{m=1}^{n} \mu_{m,n}\, \theta(s_m)\right) \exp\left\{-\frac{\beta^2}{2} \sum_{p,q=1}^{n} \mu_{p,n}\, \mu_{q,n}\, R_a(s_p - s_q)\right\} \qquad (15)$$

We thus obtain from (4), (8), (13), (14), and (15):

$$\hat{a}(t) \cong \frac{\sum_{n=1}^{\infty} \frac{1}{(2n)!} K_{2n}(t)}{1 + \sum_{n=1}^{\infty} \frac{1}{(2n)!} L_{2n}} \qquad (16)$$

where

$$K_{2n}(\tau) = -\beta\binom{2n}{n}\left(\frac{E_c^2}{2N_0}\right)^{2n}\int_0^T\cdots\int_0^T\left[\prod_{i=1}^{2n}\Xi(s_i)\,ds_i\right]\sum_{j=1}^{2n}\mu_{j,2n}R_a(s_j-\tau)\cdot$$

$$\cdot\sin\left(\sum_{m=1}^{2n}\mu_{m,2n}\theta(s_m)\right)\exp\left\{-\frac{\beta^2}{2}\sum_{p,q=1}^{2n}\mu_{p,2n}\mu_{q,2n}R_a(s_p-s_q)\right\}$$

$$L_{2n} = \binom{2n}{n}\left(\frac{E_c^2}{2N_0}\right)^{2n}\int_0^T\cdots\int_0^T\left[\prod_{i=1}^{2n}\Xi(s_i)\,ds_i\right]\cos\left(\sum_{m=1}^{2n}\mu_{m,2n}\theta(s_m)\right)\cdot$$

$$\cdot\exp\left\{-\frac{\beta^2}{2}\sum_{p,q=1}^{2n}\mu_{p,2n}\mu_{q,2n}R_a(s_p-s_q)\right\}$$

In particular

$$K_2(\tau) = \beta\frac{E_c^2}{N_0^2}\int_0^T ds_1\int_0^T ds_2\left[\gamma_0(s_1)\gamma_1(s_2)-\gamma_1(s_1)\gamma_0(s_2)\right]R_a(s_1-\tau)e^{-\beta^2(1-R_a(s_1-s_2))}$$

$$L_2 = \frac{E_c^2}{2N_0^2}\int_0^T ds_1\int_0^T ds_2\left[\gamma_0(s_1)\gamma_0(s_2)+\gamma_1(s_1)\gamma_1(s_2)\right]e^{-\beta^2(1-R_a(s_1-s_2))}$$

The above solution is based on a random carrier initial phase Φ. On the other hand, if Φ were known precisely at the receiver, then by appropriate modification of (13) it is not hard to verify that the solution becomes

$$\hat{a}(\tau) = \frac{\sum_{n=1}^{\infty}\frac{1}{n!}K_n(\tau)}{1+\sum_{n=1}^{\infty}\frac{1}{n!}L_n}\;;\quad 0\le\tau\le T \qquad (17)$$

where

$$\overline{K_n(\tau)} = \beta\left(\frac{E_0^2}{2N_0}\right)^n \int_0^T \cdots \int_0^T \left[\prod_{i=1}^n x(s_i)\,ds_i\right] \sum_{j=0}^n \left[\binom{n}{j} \sum_{\ell=1}^n \sigma_{\ell jn} R_a(s_\ell - \tau) \cdot \right.$$

$$\left. \cdot \sin\left(\sum_{m=1}^n \sigma_{mjn}(\Phi + \theta(s_m))\right) \cdot \exp\left\{-\frac{\beta^2}{2} \sum_{p,q=1}^n \sigma_{pjn}\sigma_{qjn} R_a(s_p - s_q)\right\}\right]$$

$$\overline{L_n} = \left(\frac{E_0^2}{2N_0}\right)^n \int_0^T \cdots \int_0^T \left[\prod_{i=1}^n x(s_i)\,ds_i\right] \sum_{j=0}^n \left[\binom{n}{j} \cos\left(\sum_{m=1}^n \sigma_{mjn}(\Phi + \theta(s_m))\right) \cdot \right.$$

$$\left. \cdot \exp\left\{-\frac{\beta^2}{2} \sum_{p,q=1}^n \sigma_{pjn}\sigma_{qjn} R_a(s_p - s_q)\right\}\right]$$

$$\sigma_{mjn} = \begin{cases} 1 & ; \ 1 \le m \le j \\ -1 & ; \ j < m \le n \end{cases}$$

As a final note to this section, we point out that the stationarity requirement on a(t) is not essential, but imposed only to simplify the present development. It can be shown that, if stationarity is not assumed, then (with the covariance R_a now a function of two variables) all of the above results continue to hold with the exception that for each appearance of R_a the minus sign in the argument is replaced by a comma.

Extension to Generalized Frequency - Phase Modulation

We replace (1) by

$$e_R(t) = E_o \cos\left[\omega_o t + \Phi + \beta b(t)\right] + n(t) \tag{18}$$

where

$$b(t) = (1+\gamma)\int_{-\infty}^{t} e^{-\gamma(t-\tau)} a(\tau)\, d\tau, \quad \gamma > 0$$

all other quantities being as previously defined. Note that

$$a(t) = \frac{b'(t) + \gamma b(t)}{1+\gamma}$$

where $b'(t)$ is the derivative of $b(t)$ If we let $\gamma \to \infty$ then (18) reduces back to (1). The special case of $\gamma \to 0$ corresponds to FM. Positive γ represent intermediate forms between PM and FM.

It is well known that b(t) is also a zero-mean stationary Gaussian process with covariance function R_b defined by:

$$R_b(\tau) = \frac{(1+\gamma)^2}{2\gamma}\int_{-\infty}^{\infty} R_a(\sigma) e^{-\gamma|\sigma-\tau|}\, d\sigma$$

Then by (2) we may write the MMSE estimate of $b(\tau)$ given e_R as

$$\hat{b}(\tau) = \left\langle b(\tau) \middle| e_R \right\rangle$$

and the series solutions (16), (17) with R_a replaced by R_b represents $\hat{b}(\tau)$; $0 \le \tau \le T$ Furthermore, it can be shown that

$$\hat{a}(\tau) = \left\langle a(\tau) \middle| e_R \right\rangle = \left\langle \frac{b'(\tau) + \gamma b(\tau)}{1+\gamma} \middle| e_R \right\rangle = \left(\frac{\frac{d}{d\tau} + \gamma}{1+\gamma}\right)\hat{b}(\tau) \tag{19}$$

Thus, (19) together with the explicit solution for $\hat{b}(t)$ given above under replacement of R_a by R_b describes MMSE demodulation of FM-PM within the measurement interval [0, T]. This result is a special case of a more general theorem, namely, that if (18) holds and b = La where L is a time-invariant, invertible, linear transformation (and a is as previously defined), then

$$\hat{a}(t) = (L^{-1}\hat{b})(t) \; ; \; 0 \le t \le T$$

3. ASYMPTOTIC OUTPUT SIGNAL-TO-NOISE RATIO AT LOW INPUT SIGNAL-TO-NOISE RATIOS

Phase Modulation

For very small input (RF) signal-to-noise ratios the MMSE estimates of the modulation, Equations (17) and (16), reduce (with high probability)* to the approximate forms:

$$\hat{a}(t) = \begin{cases} -\beta \frac{E_0}{N_0} e^{-\frac{1}{2}\beta^2 \int_0^T y_1(s) R_a(s-t)\, ds} \; ; & \Phi \equiv 0 \\ \frac{\beta}{2} \frac{E_0^2}{N_0^2} \int_0^T \int_0^T \left[y_c(s_1) y_1(s_2) - y_1(s_1) y_0(s_2) \right] R_a(s_1 - t)\, e^{-\beta^2 (1 - R_a(s_1 - s_2))}\, ds_1 ds_2 ; & \\ & \Phi \text{ uniform on } [0, 2\pi] \end{cases}$$

---

*Note that $|K_{2n}| \le \beta \binom{2n}{n} \left(\frac{E_0^2}{2N_0}\right)^{2n} \cdot 2n\, T^{2n} M^{2n} < \beta \left(\frac{E_0^2 T}{N_0} \cdot M\right)^{2n} \cdot 2n$ where $M = \max_{0 \le t \le T} \left(\dot{x}(t)\right)$ Hence, approximation by leading term becomes valid for a small input signal-to-noise ratio and given received signal, viz., when $E_0^2 T / N_0 << M^{-1}$. One should be careful to note that the bound M is dependent on the particular signal received and is in fact itself unbounded over the set of all possible signals. However, it is not difficult to show that, for sufficiently large $\tilde{M}$, $P\{M < \tilde{M}\}$ can be made as close to unity as one wishes. Similar reasoning applies to L_{2n}, K_n, L_n. Hence, it is necessary to qualify the above approximation by "with high probability".

From the definition of y_0, y_1 we may represent these as

$$y_0(t) = E_0 \cos\left(\beta a(t) + \Phi\right) + n_c(t)$$
$$y_1(t) = E_0 \sin\left(\beta a(t) + \Phi\right) + n_s(t)$$

where n_c and n_s are independent baseband Gaussian processes each with spectral density $2N_0$.

In view of the nonlinear operations required for MMSE estimation of angle modulated signals there is no uniquely definable output signal-to-noise ratio. However, we shall adopt, as a natural extension from the linear case, the definition of output signal-to-noise ratio given in Chapter I, viz.,

$$\Lambda_0 = \frac{\rho^2}{1-\rho^2}$$

where

$$\rho = \frac{\langle \hat{a}(\tau)\, a(\tau)\rangle}{\left(\sigma_a^2 \cdot \sigma_{\hat{a}}^2\right)^{1/2}}$$

(The correlation between the true value of the modulation and its estimate)

Note, furthermore, a property of the MMSE estimate $\hat{a}(\tau)$, that the error must be orthogonal to the estimate, i.e.,

$$\left\langle \left(a(\tau) - \hat{a}(\tau)\right) \hat{a}(\tau) \right\rangle = 0$$

Hence

$$\left\langle \hat{a}(\tau)\, a(\tau)\right\rangle = \left\langle \hat{a}^2(\tau)\right\rangle$$

and

$$\rho^2 = \frac{\left\langle \hat{a}(\tau)\, a(\tau)\right\rangle}{\sigma_a^2} = \frac{\sigma_{\hat{a}}^2}{\sigma_a^2}$$

We proceed to compute Λ_0 for the above two asymptotic cases:

1. Clearly, $\sigma_a^2 \equiv 1$

2.
$$\left\langle \hat{a}(\tau)\, a(\tau)\right\rangle = \begin{cases} -\beta \frac{E_0}{N_0} e^{\frac{1}{2}\beta^2} \int_0^T \left\langle y_1(s)\, a(\tau)\right\rangle_{a,n} R_a(s-\tau)\, ds \;;\; \Phi \equiv 0 \\ \frac{\beta}{2} \frac{E_0^2}{N_0^2} \int_0^T \int_0^T \left\langle \left(y_0(s_1) y_1(s_2) - y_1(s_1) y_0(s_2)\right) a(\tau)\right\rangle_{a,\Phi,n} \\ \qquad \cdot R_a(s_1-\tau)\, e^{-\beta^2\left(1-\delta_a(s_1-s_2)\right)} ds_1 ds_2 \;;\; \Phi \text{ uniform on } [0, 2\pi] \end{cases}$$

$$\left\langle y_1(s)\, a(\tau)\right\rangle_{a,n} = -E_0\left\langle a(\tau)\sin\beta a(s)\right\rangle_a \qquad \text{(by virtue of a, n independence and } \Phi \equiv 0\text{)}$$

$$\doteq -\beta E_0\, e^{-\frac{1}{2}\beta^2} R_a(s-\tau) \qquad \text{(See Appendix A)}$$

$$\left\langle \left(y_0(s_1)\, y_1(s_2) - y_1(s_1)\, y_0(s_2)\right) a(\tau)\right\rangle_{a,\Phi,n}$$

$$= E_0^2\left\langle \left(\sin\left(\beta a(s_1)+\Phi\right)\cos\left(\beta a(s_2)+\Phi\right)\right.\right.$$

$$\left.\left. - \cos\left(\beta a(s_1)+\Phi\right)\sin\left(\beta a(s_2)+\Phi\right)\right) a(\tau)\right\rangle_{a,\Phi}$$

(Again by a, n independence)

$$\doteq E_0^2\left\langle a(\tau)\sin\left(\beta\left(a(s_1)-a(s_2)\right)\right)\right\rangle_a$$

$$\doteq \beta E_0^2\, e^{-\beta^2\left(1-R_a(s_1-s_2)\right)}\left[R_a(s_1-\tau)-R_a(s_2-\tau)\right]$$

(See Appendix B)

Hence, for sufficiently small input signal-to-noise ratios (as measured by E_0^2/N_0) the output signal-to-noise ratio is approximated by:

$$\Lambda_0 \cong \left\langle \hat{a}(\tau)\, a(\tau)\right\rangle = \begin{cases} \left[\beta^2 e^{-\beta^2}\int_0^T R_a^2(s-\tau)\,ds\right]\dfrac{E_0^2}{N_0}\ ; & \Phi \doteq 0 \\ \left[\dfrac{\beta^2}{2}\int_0^T\int_0^T R_a(s_1-\tau)\left[R_a(s_1-\tau)-R_a(s_2-\tau)\right]\cdot\right. & \\ \left.\cdot\, e^{-2\beta^2\left(1-R_a(s_1-s_2)\right)}\,ds_1\,ds_2\right]\left(\dfrac{E_0^2}{N_0}\right)^2; & \\ \qquad\qquad \Phi \text{ uniform on } [0,2\pi] & \end{cases}$$

$$= \begin{cases} A\,\dfrac{E_0^2}{N_0}\ ; & \Phi \equiv 0 \\ B\left(\dfrac{E_0^2}{N_0}\right)^2\ ; & \Phi \text{ uniform on } [0,2\pi] \end{cases}$$

It is of interest to note (under the reasonable condition: $\int_0^\infty |R_a(t)|^{2-\delta}\, dt < \infty$ for some $\delta \geq 0$) that as $T \to \infty$ A remains bounded whereas $B \to \infty$ (see Appendix C). Since knowledge of the transmitted reference phase cannot result in a worsening of the estimate, we can be certain that $\Lambda_0(\Phi \text{ uniform on } [0, 2\pi]) \leq \Lambda_0(\Phi \equiv 0)$. Hence, the validity of the asymptotic result shown for the case of Φ uniform on $[0, 2\pi]$ must be restricted to ever smaller input E_0^2/N_0 as the observation interval is lengthened. An explanation for this effect is that whereas T's much longer than the modulation correlation time provide essentially no improvement for the case of $\Phi \equiv 0$, if on the other hand Φ is a priori uniform on $[0, 2\pi]$ then further information can be gained by very long T, namely, an auxiliary estimate of Φ which approaches the situation of known, fixed Φ as $T \to \infty$. For this reason, if we restrict our attention to the case of very large observation intervals, an adequate representation of asymptotic behavior of MMSE phase demodulation at low input E_0^2/N_0 is given by the result for $\Phi \equiv 0$. Explicitly, for two special values of τ:

$$\Lambda_0 \cong \begin{cases} \left[\beta^2 e^{-\beta^2 \int_0^\infty R_a^2(s)\, ds}\right] \frac{E_0^2}{N_0}\,; & \text{zero delay } (\tau = T) \\ \left[2\beta^2 e^{-\beta^2 \int_0^\infty R_a^2(s)\, ds}\right] \frac{E_0^2}{N_0}\,; & \text{infinite delay } \left(\tau = \frac{T}{2}\right) \end{cases} \tag{20}$$

Note further that in this regime Λ_0 may be maximized with respect to modulation index by choice of $\beta = 1$, viz.,

$$\max_{\beta}(\Lambda_0) = \begin{cases} \left[e^{-1} \int_0^\infty R_a^2(s)\, ds\right] \frac{E_0^2}{N_0}\,; & \text{zero delay} \\ \left[2e^{-1} \int_0^\infty R_a^2(s)\, ds\right] \frac{E_0^2}{N_0}\,; & \text{infinite delay} \end{cases} \tag{21}$$

For the special case of modulation a(t) having a flat spectrum out to W (Hz) and zero beyond, (20) becomes

$$\Lambda_0 \cong \begin{cases} \frac{\beta^2}{2} e^{-\beta^2} \frac{E_0^2}{2WN_0} = \frac{\beta^2}{2} e^{-\beta^2} (S/N)_i ; & \text{zero delay} \\ \beta^2 e^{-\beta^2} \frac{E_0^2}{2WN_0} = \beta^2 e^{-\beta^2} (S/N)_i ; & \text{infinite delay} \end{cases} \tag{22}$$

where $(S/N)_i \triangleq \frac{E_0^2}{2WN_0}$ the intrinsic signal-to-noise ratio.

Note, for later comparison with FM results that, in view of the normalized definition of a(t), β is also the standard deviation of the carrier phase due to modulation.

Comparison with Equation (3) of Chapter IV of this report shows that the asymptotic behavior of Λ_0 as $(S/N)_i \longrightarrow \infty$ is given (for infinite delay case) by

$$\Lambda_0 \cong \beta^2 (S/N)_i$$

Figure 1 presents plots of $\Lambda_0 / (S/N)_i$ vs β for both of these cases. The difference between these curves is an indication of the severity of the threshold. It will be noted that the high and the low $(S/N)_i$ asymptotes coincide for $\beta << 1$. The reason for this is that with $\beta << 1$ phase modulation reduces to a linear process and it is shown in Chapter IV, Appendix C of this report that for such a linear system $\Lambda_0 = \beta^2 (S/N)_i$

Frequency-Phase Modulation

Consider now the case where the modulation a(t) is pre-emphasized by a time-invariant, invertible, linear transformation L to yield b(t) which then serves as the phase modulation for transmission. At the receiving end a MMSE estimate $\hat{b}(t)$ is made and this is then subject to the inverse L^{-1} of the original transformation to yield $\hat{a}(t)$, the MMSE estimate of a(t) (see Section 2). Our particular interest is the family of transformations L_γ defined (for $\gamma > 0$) by:

$$(L_\gamma a)(t) = (1+\gamma) \int_{-\infty}^{t} e^{-\gamma(t-\tau)} a(\tau) \, d\tau$$

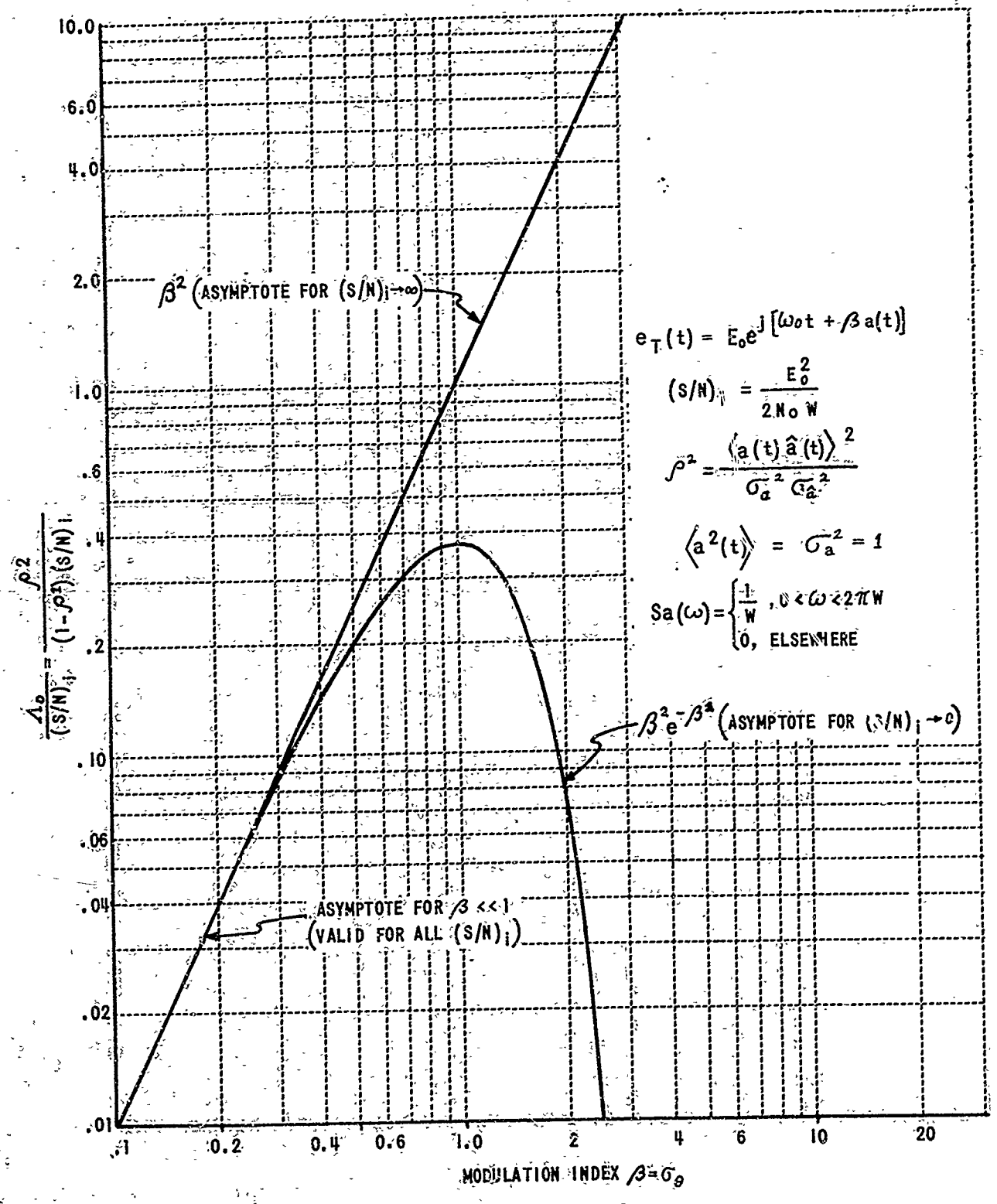


Figure III-1 PHASE MODULATION, INFINITE DELAY.

Recall that the two limits (i) $\gamma \longrightarrow \infty$ and (ii) $\gamma \longrightarrow 0$ correspond to pure PM and pure FM respectively.

Returning to the general transformation L, we repeat the calculation of

$$\Lambda_0 = \frac{\rho^2}{1-\rho^2} \cong \rho^2 = \frac{\left\langle \hat{a}(t)\, a(\tau) \right\rangle^2}{\sigma_a^2 \cdot \sigma_{\hat{a}}^2} \cong \frac{\left\langle \hat{a}(\tau)\, a(\tau) \right\rangle}{\sigma_a^{-2}}$$

$$= \left\langle \hat{a}(\tau)\, a(\tau) \right\rangle$$

as previously described for pure PM, but, in the light of earlier remarks, only for the case $\Phi = 0$. Some of the steps which follow require detailed although straightforward justification which has not been included.

$$\left\langle \hat{a}(\tau)\, a(\tau) \right\rangle = \left\langle \left(L^{-1}\hat{b}\right)(\tau) \left(L^{-1} b\right)(\tau) \right\rangle$$

$$= \left(L^{-1}(u) \left(L^{-1}(v) \left\langle \hat{b}(u)\, b(v) \right\rangle \right)(\tau) \right)(\tau) \quad *$$

$$= \left(L^{-1}(u) \left(L^{-1}(v) \left(-\beta \frac{E_0}{N_0}\, e^{-\frac{1}{2}\beta^2\sigma_b^2} \int_0^T \left\langle y_1(s) b(v) \right\rangle R_b(s-u)\, ds \right)(\tau) \right)(\tau) \right)(\tau)$$

$$= \beta^2 \frac{E_0^2}{N_0}\, e^{-\beta^2\sigma_b^2} \left(L^{-1}(u) \left(L^{-1}(v) \int_0^T R_b(s-u) R_b(s-v)\, ds \right)(\tau) \right)(\tau)$$

---

*The use of dummy variables u, v serves to indicate the scope of each L^{-1} operator.

where $\sigma_b^2 = R_b(0)$, the variance of $b(t)$.

But,

$$R_b(s-u) = \langle b(s)\, b(u) \rangle = \langle (La)(s)\,(La)(u) \rangle$$

$$= \left(L(\mu) \left(L(\nu) \langle a(\mu)\, a(\nu) \rangle \right)(s) \right)(u)$$

$$= \left(L(\mu) \left(L(\nu) R_a(\mu-\nu) \right)(s) \right)(u)$$

Hence,

$$\Lambda_0 \cong \left[\beta^2 e^{-\beta^2 \sigma_b^2} \int_0^T \left(L(\nu) R_a(\tau-\nu) \right)^2 (s)\, ds \right] \frac{E_0^2}{N_0}$$

where

$$\sigma_b^2 = R_b(0) = \left(L(\mu) \left(L(\nu) R_a(\mu-\nu) \right)(0) \right)(0)$$

Specializing L to the form L_γ and assuming infinite observation time $T \to \infty$ with two subcases of zero delay ($\tau = T$) and infinite delay $\tau = \frac{1}{2}T$, we obtain:

$$\Lambda_0 \cong \begin{cases} \left[\beta^2 (1+\gamma)^2 e^{-\beta \sigma_b^2} \int_0^\infty \left(\int_s^\infty e^{-\gamma(\nu-s)} R_a(\nu)\, d\nu \right)^2 ds \right] \frac{E_0^2}{N_0} ; & \text{zero delay} \\ \left[\beta^2 (1+\gamma)^2 e^{-\beta^2 \sigma_b^2} \int_{-\infty}^\infty \left(\int_s^\infty e^{-\gamma(\nu-s)} R_a(\nu)\, d\nu \right)^2 ds \right] \frac{E_0^2}{N_0} ; & \text{infinite delay} \end{cases}$$

where

$$\sigma_b^2 = \frac{(1+\gamma)^2}{\gamma} \int_0^\infty e^{-\gamma y} R_a(y)\, dy$$

(23)

We specialize further to the limiting case $\gamma \to 0$, corresponding to pure FM, and obtain:

$$\Lambda_0 \cong \lim_{\gamma \to 0} \left[\frac{\beta^2}{\gamma} \exp\left\{ -\frac{\beta^2}{\gamma} \int_0^\infty e^{-\gamma y} R_a(y)\, dy \right\} \cdot \gamma \int_{0 \text{ or } -\infty} \left(\int_s^\infty e^{-\gamma(v-s)} R_a(v)\, dv \right)^2 ds \right] \frac{E_0^2}{N_0}$$

Note that for this limiting case to make sense σ_b^2 must remain finite (otherwise b(t) does not exist); hence we require:

$$0 < \lim_{\gamma \to 0} \left[\frac{1}{\gamma} \int_0^\infty e^{-\gamma y} R_a(y)\, dy \right] < \infty$$

(This implies that the DC power spectral density of a(t) must be zero.)

The above restriction on a(t) is of course not an essential limitation in actual FM; it arises here because of the need to keep b(t) stationary so that the prior stationary analysis could be applied. It would have been eliminated, at the cost of additional complexity, if the more general nonstationary analysis had been employed.

By reversal of order of integration, the last integral factor (along with γ coefficient) in the expression for Λ_0 can readily be transformed to:

$$\tfrac{1}{2} \int_0^\infty d\mu \int_0^\infty dv \left[e^{-\gamma|\mu - v|} - e^{-\gamma(\mu+v)} \right] R_a(\mu) R_a(v) \quad ; \text{ zero delay}$$

$$\tfrac{1}{2} \int_{-\infty}^\infty d\mu \int_{-\infty}^\infty dv\, e^{-\gamma|\mu - v|} R_a(\mu) R_a(v) \quad ; \quad \text{infinite delay}$$

Hence,

$$\Lambda_0 \cong \begin{cases} \left[\dfrac{\beta^2}{2} \exp\left\{ -\beta^2 \lim\limits_{\gamma \to 0} \dfrac{1}{\gamma} \int_0^\infty e^{-\gamma y} R_a(y)\, dy \right\} \cdot \right. \\ \qquad \left. \cdot \lim\limits_{\gamma' \to 0} \dfrac{1}{\gamma'} \int_0^\infty d\mu \int_0^\infty dv \left[e^{-\gamma'|\mu - v|} - e^{-\gamma'(\mu+v)} \right] R_a(\mu) R_a(v) \right] \dfrac{E_0^2}{N_0} \text{ ; zero delay} \\ \left[\dfrac{\beta^2}{2} \exp\left\{ -\beta^2 \lim\limits_{\gamma \to 0} \dfrac{1}{\gamma} \int_0^\infty e^{-\gamma y} R_a(y)\, dy \right\} \right. \\ \qquad \left. \cdot \lim\limits_{\gamma' \to 0} \dfrac{1}{\gamma'} \int_{-\infty}^\infty d\mu \int_{-\infty}^\infty dv\, e^{-\gamma'|\mu - v|} R_a(\mu) R_a(v) \right] \dfrac{E_0^2}{N_0} \text{ ; infinite delay} \end{cases} \tag{24}$$

Again, as in the PM case, Λ_0 can be maximized with respect to β by requiring that the exponent of e be -1. This leads to:

$$\max_{\beta}(\Lambda_0) \cong \begin{cases} \left[\frac{1}{2e}\,\frac{\lim_{\gamma\to 0}\frac{1}{\gamma}\int_0^\infty d\mu\int_0^\infty d\nu\left[e^{-\gamma|\mu-\nu|}-e^{-\gamma(\mu+\nu)}\right]R_a(\mu)R_a(\nu)}{\lim_{\gamma'\to 0}\frac{1}{\gamma'}\int_0^\infty e^{-\gamma' y}R_a(y)\,dy}\right]\frac{E_0^2}{N_0}\ ; \text{zero delay} \\ \left[\frac{1}{2e}\,\frac{\lim_{\gamma\to 0}\frac{1}{\gamma}\int_{-\infty}^\infty d\mu\int_{-\infty}^\infty d\mu\ e^{-\gamma|\mu-\nu|}R_a(\mu)R_a(\nu)}{\lim_{\gamma'\to 0}\frac{1}{\gamma'}\int_0^\infty e^{-\gamma' y}R_a(y)\,dy}\right]\frac{E_c^2}{N_0}\ ; \text{infinite delay} \end{cases} \tag{25}$$

and

$$\beta^2_{MAX} = \left[\lim_{\gamma\to 0}\frac{1}{\gamma}\int_0^\infty e^{-\gamma y}R_a(y)\,dy\right]^{-1}$$

Application to Modulation Functions Having a Bandlimited White Spectrum and Comparison with Large-Signal Asymptotic Result

Consider the special case of a(t) with bandlimited white spectrum within $[\,W_1\,,\,W_2\,]$. We then have

$$R_a(t) = \frac{\sin 2\pi W_2 t - \sin 2\pi W_1 t}{2\pi(W_2-W_1)t} \tag{25a}$$

Equation (24) then yields for the infinite delay case (See Appendix C):

$$\Lambda_0 \cong \left[\frac{\beta^2}{4\pi^2 W_1 W_2 (W_2-W_1)}\ e^{-\frac{\beta^2}{4\pi W_1 W_2}}\right]\frac{E_0^2}{N_0}$$

$$\cong \sigma_\theta^2\, e^{-\sigma_\theta^2}(S/N)_i \quad ;\ \text{FM, infinite delay, } (S/N)_i \ll 1 \tag{26}$$

where

$$\sigma_\theta = \frac{\beta}{2\pi\sqrt{W_1 W_2}} \quad ,$$ the standard deviation of the phase angle due to modulation (See Equations (18), (23), and first part of Appendix C).

$$(S/N)_i = \frac{E_o^2}{2(W_2 - W_1)N_o} \quad ,$$ intrinsic signal-to-noise ratio

Note the similarity to Equation (22) for the PM case.

For the case of large $(S/N)_i$ (and infinite delay) we compute the output noise power H_∞ to be (see Appendix E)

$$H_\infty \cong \frac{8\pi^2(W_2^3 - W_1^3)}{3\beta^2} \cdot \frac{N_o}{E_o^2} = \frac{8\pi^2 W_1 W_2 \left(\frac{W_2}{W_1} + 1 + \frac{W_1}{W_2}\right)(W_2 - W_1)N_o}{3\beta^2 E_o^2}$$

Hence the output signal-to-noise ratio is

$$\Lambda_o = \frac{\rho^2}{1-\rho^2} = \frac{\langle \hat{a}^2 \rangle}{\langle a^2 \rangle - \langle \hat{a}^2 \rangle} \cong \frac{\langle a^2 \rangle}{H_\infty} = H_\infty^{-1}$$

$$= \frac{3}{\frac{W_2}{W_1} + 1 + \frac{W_1}{W_2}} \cdot \sigma_\theta^2 \, (S/N)_i \; ; \; \text{FM, infinite delay, } (S/N)_i >> 1 \qquad (27)$$

In Chapter IV, Appendix C of this report a linear analysis, appropriate for $\sigma_\theta << 1$, yields FM output signal-to-noise ratio for bandlimited a(t) valid for all $(S/N)_i$. As would be expected this general linear result reduces to Equations (26) and (27) for $(S/N)_i << 1$ and $(S/N)_i >> 1$ respectively. However, except for $(S/N)_i >> 1$, linear analysis loses all validity, for larger σ_θ.

Figure 2 provides plots of $\Lambda_o / (S/N)_i$ vs σ_θ for the two asymptotic cases, Equations (26) and (27). For large $\sigma_\theta = \beta/2\pi\sqrt{W_1 W_2}$ and small or moderate $\beta = W_2/W_1$ the difference between these two values (the asymptotic intercepts on a log-log plot of Λ_o vs $(S/N)_i$) is an indication of the severity of threshold. Note that for smaller σ_θ if β is sufficiently large, not only is thresholding apparently absent but actually a reverse relation between asymptotic intercept occurs. This phenomenon is depicted in more detail in Figure 3 which shows the transitional region of intermediate $(S/N)_i$ based on the linear analysis for $\sigma_\theta << 1$

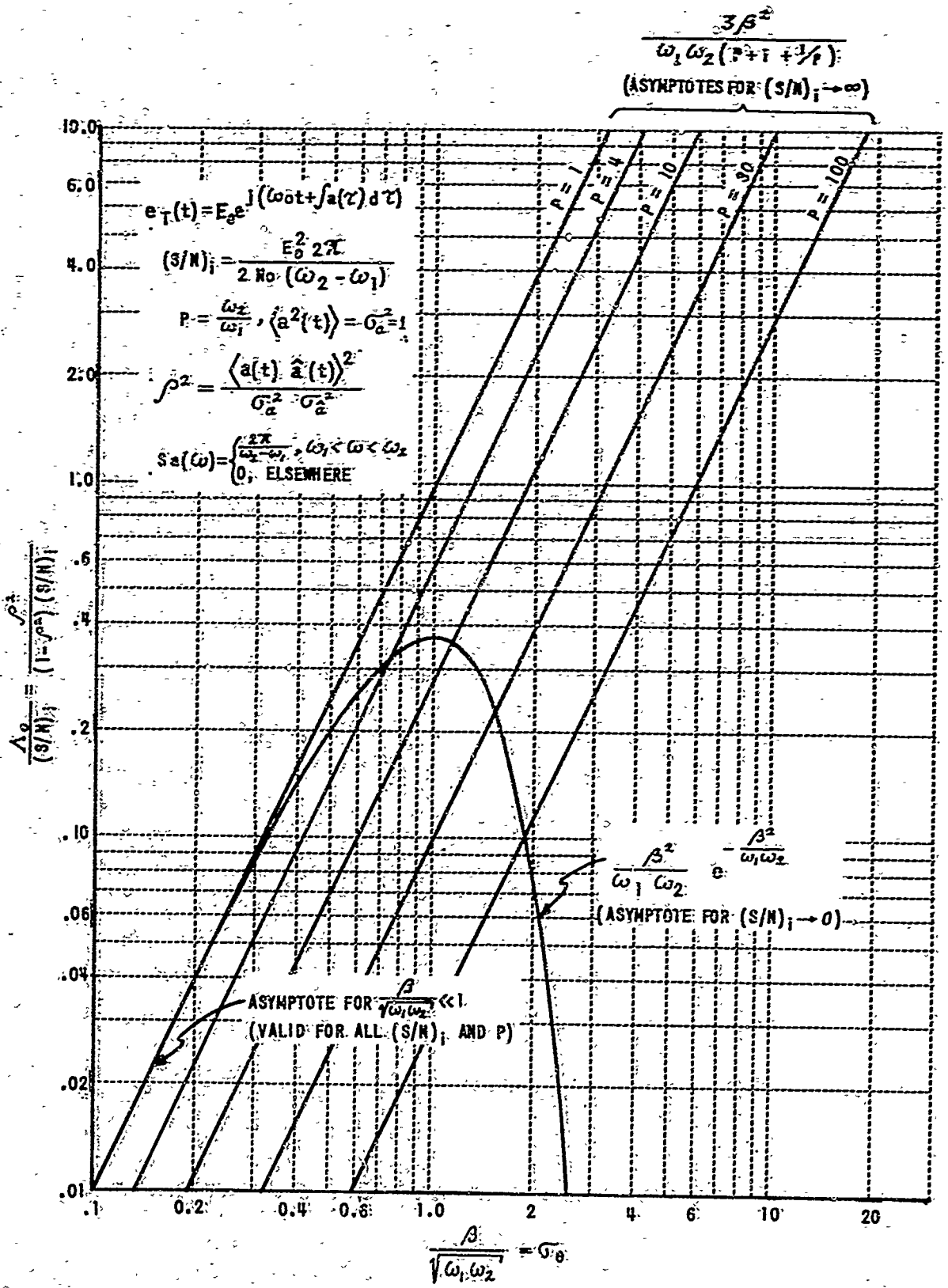


Figure III-2 FREQUENCY MODULATION, INFINITE DELAY.

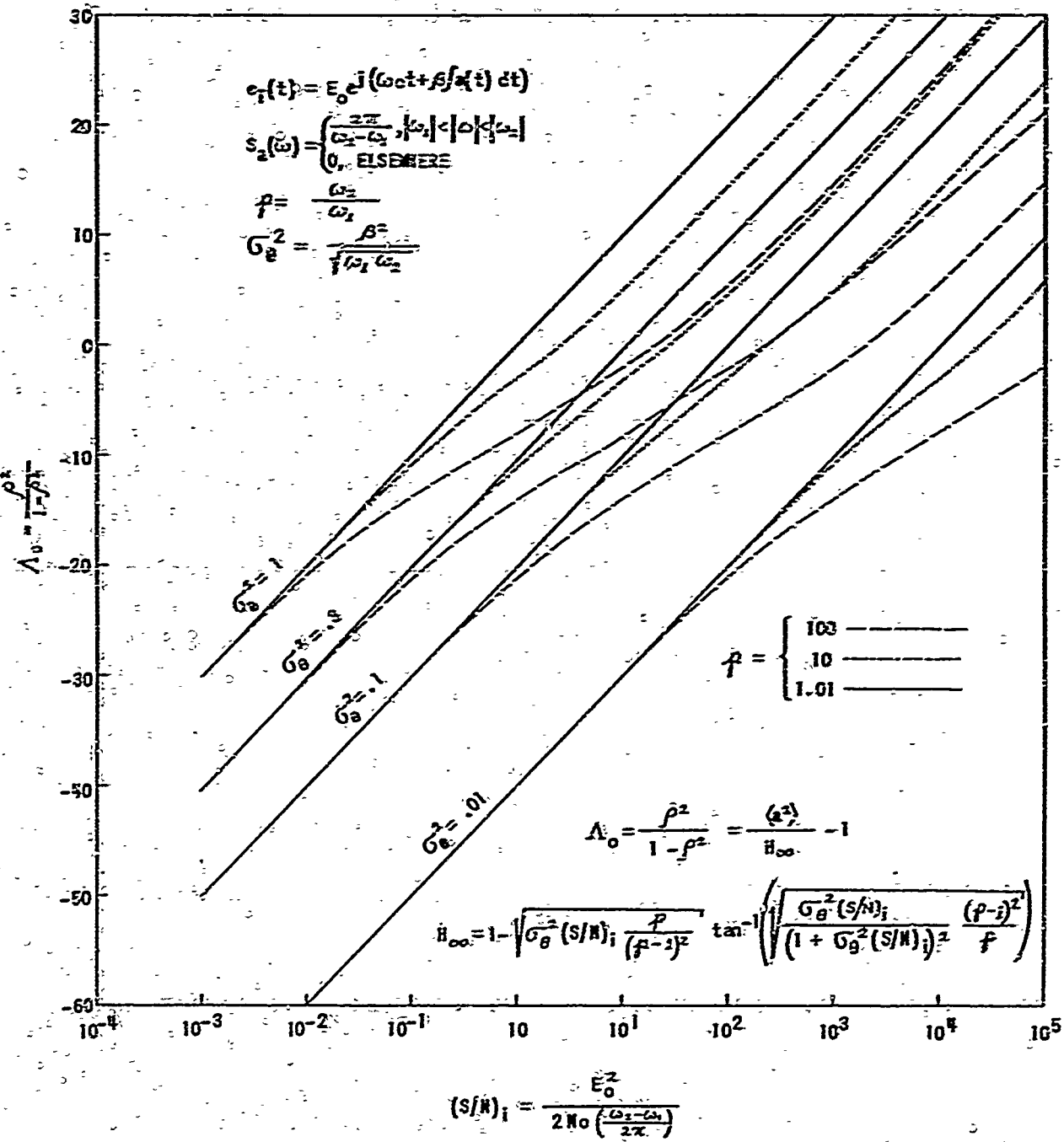


Figure III-3 Λ_0 vs $(S/N)_i$ (FROM LINEAR ANALYSIS)

given in Chapter IV, Appendix C of this report. The curves for $\sigma_\theta = .01$, .1 and .3 are reasonably accurate, but that for $\sigma_\theta = 1$ is already too high by a factor of about e at the left end. Unfortunately, these curves give no quantitative indication of the threshold type of transition at intermediate $(S/N)_i$ for larger values of σ_θ.

4. CONCLUSIONS

An analysis of angle modulated systems operating under conditions of very low input signal-to-noise ratios has been carried out. Using $\lambda_0 = \frac{\rho^2}{1-\rho^2}$ as the measure of performance, asymptotic expressions, Equations (20) and (24) for λ_0 in terms of $(S/N)_i$ and the modulation index, were developed for both PM and FM. It is interesting that λ_0 varies linearly with $(S/N)_i$ at very low $(S/N)_i$. It is of course well known, viz Chapters II, IV, and V, that λ_0 varies linearly with $(S/N)_i$ well above threshold, when the modulation spectrum is strictly bandlimited. The results here obtained show that this is true for all modulation spectra at very small $(S/N)_i$.

It has been observed (Chapters II, IV, V) that the severity of the threshold increases for both PM and FM as the modulation index is increased. Our results confirm this and show that the ratio between the large and small $(S/N)_i$ asymptotes of λ_0 vs $(S/N)_i$ increases exponentially with the square of the modulation index, (Equations (22) and (25)). The results here derived are in agreement with the linearized analysis (Appendix IV-C), which is asymptotically valid at all $(S/N)_i$ as the modulation index goes to zero.

APPENDIX III-A

The pair $(\beta a(\tau), \beta a(s))$ is bivariate normal with zero mean and covariance matrix

$$\begin{pmatrix} \beta^2 & \beta^2 R_a(s-\tau) \\ \beta^2 R_a(s-\tau) & \beta^2 \end{pmatrix}.$$

Hence,

$$\left\langle \beta a(\tau) \sin \beta a(s) \right\rangle_a = \frac{1}{2\pi\beta^2\sqrt{1-R_a^2(s-\tau)}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \sin y \; e^{-\frac{1}{2}\left[\frac{x^2+y^2-2xy\,R_a(s-\tau)}{\beta^2(1-R_a^2(s-\tau))}\right]} dx\, dy.$$

$$= \frac{R_a(s-\tau)}{\sqrt{2\pi}\,\beta} \int_{-\infty}^{\infty} y \sin y \; e^{-\frac{1}{2}\frac{y^2}{\beta^2}} dy$$

$$= \frac{\beta^2 R_a(s-\tau)}{\sqrt{2\pi}\,\beta} \int_{-\infty}^{\infty} \cos y \; e^{-\frac{1}{2}\frac{y^2}{\beta^2}} dy$$

$$= \beta^2 R_a(s-\tau)\, e^{-\frac{1}{2}\beta^2}$$

APPENDIX III-B

The pair $\{a(\tau), \beta(a(s_1)-a(s_2))\}$ is bivariate normal with zero mean and covariance matrix

$$\begin{pmatrix} \beta^2 & \beta\left(R_a(s_1-\tau)-R_a(s_2-\tau)\right) \\ \beta\left(R_a(s_1-\tau)-R_a(s_2-\tau)\right) & 2\beta^2(1-R_a(s_1-s_2)) \end{pmatrix}$$

Transforming from $a(\tau)$ to $\overline{a(\tau)} = \beta\sqrt{2(1-R_a(s_1-s_2))}\cdot a(\tau)$ we can immediately apply the result of Appendix A to obtain

$$\left\langle \overline{a(\tau)} \sin \beta(a(s_1)-a(s_2)) \right\rangle_a$$

$$= \sqrt{2(1-R_a(s_1-s_2))}\cdot\beta^2\left(R_a(s_1-\tau)-R_a(s_2-\tau)\right)\cdot e^{-\beta^2(1-R_a(s_1-s_2))}$$

Consequently,

$$\left\langle a(\tau) \sin \beta(a(s_1)-a(s_2)) \right\rangle_a$$

$$= \frac{\left\langle \overline{a(\tau)} \sin \beta(a(s_1)-a(s_2)) \right\rangle_a}{\beta\sqrt{2(1-R_a(s_1-s_2))}}$$

$$= \beta\left[R_a(s_1-\tau)-R_a(s_2-\tau)\right] e^{-\beta^2(1-R_a(s_1-s_2))}$$

APPENDIX III-C

We are to prove under the hypotheses:

(i) $R_a \neq 0$ on set of positive measure; $|R_a| \leq 1$

(ii) $\int_0^\infty |R_a(t)|^{2-\delta}\, dt < \infty$ for some $\delta > 0$

that

$$\lim_{T \to \infty} \int_0^T \int_0^T R_a(s_1 - \tau)\left[R_a(s_1 - \tau) - R_a(s_2 - \tau)\right] e^{-2\beta^2(1 - R_a(s_1 - s_2))}\, ds_1\, ds_2 \longrightarrow +\infty$$

The integral expression may be written as:

$$\int_0^T \int_0^T R_a^2(s_1 - \tau)\, e^{-2\beta^2(1 - R_a(s_1 - s_2))}\, ds_1\, ds_2 - \int_0^T \int_0^T R_a(s_1 - \tau)\, R_a(s_2 - \tau) e^{-2\beta^2(1 - R_a(s_1 - s_2))}\, ds_1\, ds_2$$

$$\geq e^{-4\beta^2} \int_0^T \int_0^T R_a^2(s_1 - \tau)\, ds_1\, ds_2 - \left| \int_0^T \int_0^T R_a(s_1 - \tau) R_a(s_2 - \tau)\, e^{-2\beta^2(1 - R_a(s_1 - s_2))}\, ds_1\, ds_2 \right|$$

$$\geq T e^{-4\beta^2} \int_0^T R_a(s - \tau)\, ds - \left(\int_0^T |R_a(s - \tau)|\, ds \right)^2$$

$$= T\left[e^{-4\beta^2} \int_0^T R_a^2(s - \tau)\, ds - \left(\frac{\int_0^T |R_a(s - \tau)|\, ds}{T^{1/2}} \right) \right]$$

By (i) $\int_0^\infty R_a^2(s - \tau)\, ds > 0$; therefore we are done if we can show that

$$\lim_{T \to \infty} \left(\frac{\int_0^T |R_a(s - \tau)|\, ds}{T^{1/2}} \right) = 0$$

By Hölder's inequality with $f = R_a$, $g = 1$, $p = 2-\delta$, $q = \frac{2-\delta}{1-\delta}$

$$\int_0^T |R_a(s-\tau)|\,ds \leq \left[\int_0^T |R_a(s-\tau)|^{2-\delta}\,ds\right]^{\frac{1}{2-\delta}} \left[\int_0^T 1^{\frac{2-\delta}{1-\delta}}\,ds\right]^{\frac{1-\delta}{2-\delta}}$$

$$= T^{\frac{1-\delta}{2-\delta}} \left[\int_0^T |R_a(s-\tau)|^{2-\delta}\,ds\right]^{\frac{1}{2-\delta}}$$

Hence

$$\lim_{T\to\infty} \left(\frac{\int_0^T |R_a(s-\tau)|\,ds}{T^{1/2}}\right) \leq \lim_{T\to\infty} \left(T^{-\frac{\delta}{2(2-\delta)}} \left[\int_0^T |R_a(s-\tau)|^{2-\delta}\,ds\right]^{\frac{1}{2-\delta}}\right)$$

$$= 0 \qquad \text{by (ii)}$$

APPENDIX III-D

The limits in Equation (24) [Infinite delay case] are evaluated for R specified in Equation (25a), as follows:

$$\lim_{\gamma \to 0} \frac{1}{\gamma} \int_0^\infty e^{-\gamma y} R_a(y)\, dy = \lim_{\gamma \to 0} \frac{\partial}{\partial \gamma} \int_0^\infty e^{-\gamma y} R_a(y)\, dy$$

$$= \lim_{\gamma \to 0} \int_0^\infty -y\, e^{-\gamma y} \left[\frac{\sin \omega_2 y - \sin \omega_1 y}{(\omega_2 - \omega_1) y} \right] dy$$

$$= \frac{1}{\omega_2 - \omega_1} \left[\lim_{\gamma \to 0} \int_0^\infty e^{-\gamma y} \sin \omega_1 y\, dy - \lim_{\gamma \to 0} \int_0^\infty e^{-\gamma y} \sin \omega_2 y\, dy \right]$$

$$= \frac{1}{\omega_2 - \omega_1} \left[\frac{1}{\omega_1} - \frac{1}{\omega_2} \right] = \frac{1}{\omega_1 \omega_2} = \frac{1}{4\pi^2 W_1 W_2}$$

$$\lim_{\gamma \to 0} \frac{1}{\gamma} \int_{-\infty}^{\infty} d\mu \int_{-\infty}^{\infty} d\nu\, e^{-\gamma|\mu - \nu|} R_a(\mu) R_a(\nu) = \lim_{\gamma \to 0} \frac{2}{\gamma} \int_{-\infty}^{\infty} d\mu \int_{-\infty}^{\mu} d\nu\, e^{-\gamma(\mu - \nu)} R_a(\mu) R_a(\nu)$$

$$= \lim_{\gamma \to 0} \frac{2}{\gamma} \int_{-\infty}^{\infty} d\mu\, e^{-\gamma \mu} R_a(\mu) \int_{-\infty}^{-\mu} d\nu\, e^{\gamma \nu} R_a(\nu)$$

$$= 2 \lim_{\gamma \to 0} \left[\int_{-\infty}^{\infty} d\mu\, e^{-\gamma \mu} R_a(\mu) \int_{-\infty}^{\mu} d\nu \cdot \nu\, e^{\gamma \nu} R_a(\nu) \right.$$

$$\left. - \int_{-\infty}^{\infty} d\mu \cdot \mu e^{-\gamma \mu} R_a(\mu) \int_{-\infty}^{\mu} d\nu\, e^{\gamma \nu} R_a(\nu) \right]$$

$$= 4 \lim_{\gamma \to 0} \int_{-\infty}^{\infty} d\mu\, e^{-\gamma \mu} R_a(\mu) \int_{-\infty}^{\mu} d\nu \cdot \nu\, e^{\gamma \nu} R_a(\nu)$$

(by interchange in order of integration of the 2nd term and noting that R_a is an even function)

$$= \frac{4}{\omega_2-\omega_1} \lim_{\gamma\to 0} \int_{-\infty}^{\infty} d\mu\, e^{-\gamma\mu} R_a(\mu) \int_{-\infty}^{\mu} d\nu \cdot e^{\gamma\nu} (\sin\omega_2\nu - \sin\omega_1\nu)$$

$$= \frac{4}{\omega_2-\omega_1} \lim_{\gamma\to 0} \int_{-\infty}^{\infty} d\mu\, e^{-\gamma\mu} R_a(\mu) \cdot e^{\gamma\mu} \left[\frac{\gamma\sin\omega_2\mu - \omega_2\cos\omega_2\mu}{\omega_2^2+\gamma^2} - \frac{\gamma\sin\omega_1\mu - \omega_1\cos\omega_1\mu}{\omega_1^2+\gamma^2} \right]$$

$$= \frac{4}{(\omega_2-\omega_1)^2} \int_{-\infty}^{\infty} d\mu \frac{(\sin\omega_2\mu - \sin\omega_1\mu)}{\mu} \left[\frac{\cos\omega_1\mu}{\omega_1} - \frac{\cos\omega_2\mu}{\omega_2} \right]$$

$$= \frac{8}{(\omega_2-\omega_1)^2} \int_0^{\infty} d\mu \cdot \frac{1}{2} \left[\frac{\sin(\omega_2+\omega_1)\mu + \sin(\omega_2-\omega_1)\mu}{\omega_1\mu} - \frac{\sin 2\omega_1\mu}{\omega_1\mu} \right.$$

$$\left. + \frac{\sin(\omega_1+\omega_2)\mu + \sin(\omega_1-\omega_2)\mu}{\omega_2\mu} - \frac{\sin 2\omega_2\mu}{\omega_2\mu} \right]$$

$$= \frac{8}{(\omega_2-\omega_1)^2} \cdot \left[\frac{1}{2} \left(\frac{\pi}{2\omega_1} + \frac{\pi}{2\omega_1} - \frac{\pi}{2\omega_1} + \frac{\pi}{2\omega_2} - \frac{\pi}{2\omega_2} - \frac{\pi}{2\omega_2} \right) \right]$$

$$= \frac{2\pi}{\omega_1\,\omega_2\,(\omega_2-\omega_1)} = \frac{1}{4\pi^2\, W_1\, W_2\, (W_2-W_1)}$$

APPENDIX III-E

For very large $(S/N)_i$, the output noise power spectrum of an FM (conventional or MMSE) receiver is known to be (Chapter II of Ref. [2]).

$$N(\omega) = \frac{N_0 \omega^2}{\beta^2 E_0^2}$$

The output signal spectrum is of course identical to the input, viz.,

$$S(\omega) \begin{cases} \frac{\pi}{\omega_2 - \omega_1} , & \omega_1 < |\omega| < \omega_2 \\ 0 , & \text{elsewhere} \end{cases}$$

Applying Wiener's result for mean-square error

$$H_\infty = \int_{-\infty}^{\infty} \frac{S(\omega) N(\omega)}{S(\omega) + N(\omega)} \frac{d\omega}{2\pi}$$

we obtain

$$H_\infty = \frac{1}{\omega_2 - \omega_1} \int_{\omega_1}^{\omega_2} \frac{\omega^2 d\omega}{A^2 + \omega^2} , \qquad A^2 = \frac{\pi \beta^2 E_0^2}{N_0 (\omega_2 - \omega_1)}$$

$$= 1 - \frac{A}{\omega_2 - \omega_1} \left[\tan^{-1} \frac{\omega_2}{A} - \tan^{-1} \frac{\omega_1}{A} \right]$$

$$\cong 1 - \frac{A}{\omega_2 - \omega_1} \left[\frac{\omega_2 - \omega_1}{A} - \frac{1}{3} \left[\left(\frac{\omega_2}{A} \right)^3 - \left(\frac{\omega_1}{A} \right)^3 \right] \right] ; \text{ for } \left(S/N \right)_i \gg 1 \text{ i.e., } A^2 \gg 1$$

$$\cong \frac{1}{3(\omega_2 - \omega_1)} \frac{\omega_2^3 - \omega_1^3}{A^2} = \frac{8\pi^2 (W_2^3 - W_1^3)}{\beta^2 E_0^2}$$

REFERENCES

[1] Abbate, J.W. and Schilling, D.L., "Estimation of Random Phase- and Frequency-Modulation Signals Using a Bayes Estimator", IEEE Transactions on Information Theory, Vol. IT-11, July 1965, pp. 462-463.

[2] Lawton, John G., "Investigation of Analog and Digital Communication Systems" (Phase 3 Report), CAL Final Report UA-1420-S-3, TDR-RADC-63-147, ASTIA AD 407343, May 1963.

IV. "MONTE CARLO" EVALUATION OF THE PERFORMANCE ATTAINABLE WITH PHASE MODULATED SYSTEMS

Analytical investigation of the performance of angle modulated systems, viz. Chapters II and III, is difficult and does not yield closed form expressions valid at all signal-to-noise ratios. Further, it is necessary to assume that the modulation as well as the noise processes are Gaussian in order to proceed with the analysis. Most of the earlier analyses[1, 2] were concerned with the maximum likelihood or maximum a posteriori estimation (MAP), while the analyses presented in Chapters II and III are primarily concerned with minimum mean-square error estimation (MMSE). Analytic similarities between MAP estimation and the operation of a phase-locked loop (PLL) have frequently been noted (viz. Chapters II, V and [3, 4]). However, when phase-locked loops are used the performance of the system is presented in practice by a curve of MSE (or output signal-to-noise ratio) vs input signal-to-noise ratio. Although it is known that the MAP and MMSE estimates approach each other asymptotically at large $(S/N)_i$, it has not yet been possible to determine their relative performance in and below the threshold region by analytic techniques. For the above reasons, a "Monte Carlo" evaluation of a PM system, which can be carried out at all input signal-to-noise ratios, was performed with the use of a digital computer. This evaluation yielded upper bounds on the mean-square error resulting from use of the MMSE and MAP estimators. It was found that the evaluation of the performance of an optimum PM system by the technique described below is quite straightforward. Because only limited computer time was available for this investigation the range of parameters which could be explored had to be limited. However, the techniques described can readily be extended to cover a wider range of modulation processes and parameters than reported here. In order to facilitate additional investigations, the complete computer program, written in FORTRAN IV, is reproduced in Appendix IV-A. The results obtained are in agreement with the analytical results of Chapters II and III, wherever such a comparison can be made. In addition, the curves presented give valuable insight into the behavior of angle modulated systems operating under conditions for which adequate analytical techniques are not available.

As discussed in Chapter I the performance of an analog communication system may be conveniently described by specifying the mean-square error between the original modulation a(t) and its estimate $\hat{a}(t)$ or by the quantity $\Lambda_o = \frac{\rho^2}{1-\rho^2}$ where ρ^2 is the square of the correlation coefficient of a(t) and $\hat{a}(t)$. Both of these performance measures have been computed and plots of $\left(\frac{\sigma_a^2}{MSE}\right)$ and Λ_o vs intrinsic signal-to-noise ratio,

$$\left(\frac{S}{N}\right)_i = \left\{\frac{\text{CARRIER POWER}}{\text{RF NOISE POWER IN THE MODULATION BANDWIDTH}}\right\} \qquad (1)$$

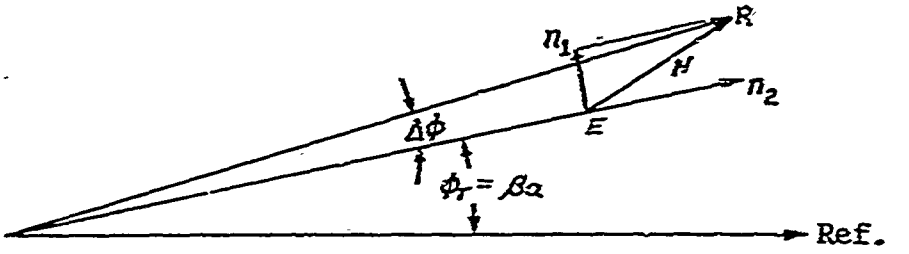
are presented in Figures IV-1, IV-2.* It will be noted that these curves are in close agreement for $(S/N)_i$ such that operation is above threshold and also into the threshold region. The behavior well below threshold differs markedly as one would expect from the discussion in Chapter I. It can also be seen that the scatter of the experimental data for small $(S/N)_i$ is much greater for Λ_o than for $\left(\frac{\sigma_a^2}{MSE}\right)$ expressed in db. The reason for this is that the evaluation of these quantities is based on a finite amount of experimental data and that at small $(S/N)_i$, Λ_o is much more sensitive to experimental errors than $\left(\frac{\sigma_a^2}{MSE}\right)$ (see Appendix IV-D). This observation should not be construed as a shortcoming of Λ_o as a measure of performance. In fact, it is the predictability (and hence the low information content) of the experimental MSE at low $(S/N)_i$ which makes it insensitive to experimental errors.

Well above threshold the RF noise in a finite bandwidth B, will with high probability have a small magnitude compared to the amplitude of the RF carrier. Under these conditions one may consider the effect of the RF noise on a phase modulated signal to be a small perturbation of the received phase angle. The asymptotic performance for large $(S/N)_i$ can then be determined by inspection of the phasor diagram.

---

*The computer printout from which Figures IV-1, 2, 6, were plotted is presented as Appendix IV-B.

The received phase angle differs from the transmitted phase angle by $\Delta\phi$. For $(S/N)_i \gg 1$, $|n_1|, |n_2| \ll E_o$ and $\Delta\phi \approx \frac{n_2}{E_o}$



n_2 is a lowpass bandlimited white Gaussian random process with bandwidth $B/2$ equal to half the RF bandwidth and power density $2N_0$ equal to twice the power density of the RF noise. The modulating signal a(t) is restricted to a bandwidth $W = B/M$, Hz where M is the bandwidth expansion factor. Errors due to components of n_2 falling outside of the band $0-W$ Hz can be eliminated by a lowpass filter. The mean-square value of n_2 over the band $0-W$ Hz is $2N_0W$. So that the mean-square error in the filtered phase angle becomes

$$\langle (\phi - \phi_T)^2 \rangle = \sigma_{\Delta\phi}^2 = \frac{2N_0W}{E_o^2} \qquad (2)$$

$$= \frac{\text{RF Noise Power in the Bandwidth of the Modulating Signal}}{\text{Carrier Power}} = (S/N)_i^{-1}$$

In Appendix IV-C it is shown that in the limit $\sigma_{\phi_T}^2 \to 0$ PM and FM may be treated as linear modulation systems at all $(S/N)_i$. For PM as $\beta \to 0$, $\Lambda_0 \to \beta^2 (S/N)_i$ for all values of $(S/N)_i$, which agrees with Equation (2) above which is valid for $(S/N)_i \to \infty$ and all values of β.

Note that Equation (2) is independent of the modulation index β or the bandwidth expansion factor $M = B/W$ However, the assumption that $|n_1|, |n_2| \ll E_o$ requires that $(S/N)_i$ be increased as M is increased.

Since the mean-square phase error* in the asymptotic region is inversely proportional to $(S/N)_i$, the abscissa of Figure IV-1 can also be labeled with $\langle(\hat{\phi}-\phi_T)^2\rangle=\sigma_{\Delta\phi}^2$ in radians^2, but this scale is applicable only in the high signal-to-noise ratio region, where the above asymptotic approximations are valid.

Since $\hat{\phi} = \beta\hat{a}$ the mean-square error in the estimate of the modulation is related to the mean-square error in the estimate of ϕ_T by

$$MSE = \left\langle(\hat{a}-a)^2\right\rangle=\frac{1}{\beta^2}\left\langle(\hat{\phi}-\phi_T)^2\right\rangle = \frac{\sigma_{\Delta\phi}^2}{\beta^2} \sim \frac{1}{\beta^2(S/N)_i} \qquad (3)$$

Therefore, the ordinate of Figure IV-1, $\frac{\sigma_a^2}{MSE} = \frac{\beta^2\sigma_a^2}{\sigma_{\Delta\phi}^2}$.

Note that this relationship is valid for all operating conditions, i.e., below as well as above threshold. The asymptotic relationship between $\frac{\sigma_a^2}{MSE}$ and the RF signal-to-noise ratio is given by

$$\frac{\sigma_a^2}{MSE} = \frac{\beta^2\sigma_a^2}{\sigma_{\Delta\phi}^2} = \frac{\beta^2\sigma_a^2 E_o^2}{2N_o W} = \beta^2\sigma_a^2(S/N)_i$$

$$= \beta^2\sigma_a^2\ \frac{B}{W}\ \frac{E_o^2}{2N_o B} = \beta^2 M\sigma_a^2(S/N)_{RF} \qquad (4)$$

Note that this result is in agreement with the asymptotic form of the results of the MMSE analysis of Chapter II.

In Figure IV-1 the above asymptotic expressions are plotted together with the experimental data. The agreement between theory and experiment

---

*The filtered phase error is a mathematical entity defined by $\langle(\hat{\phi}-\phi_T)^2\rangle = \beta^2\langle(\hat{a}-a)^2\rangle$ It may, however, be compared with the phase error existing in the low frequency equivalent of a phase-locked loop. (Chapter V).

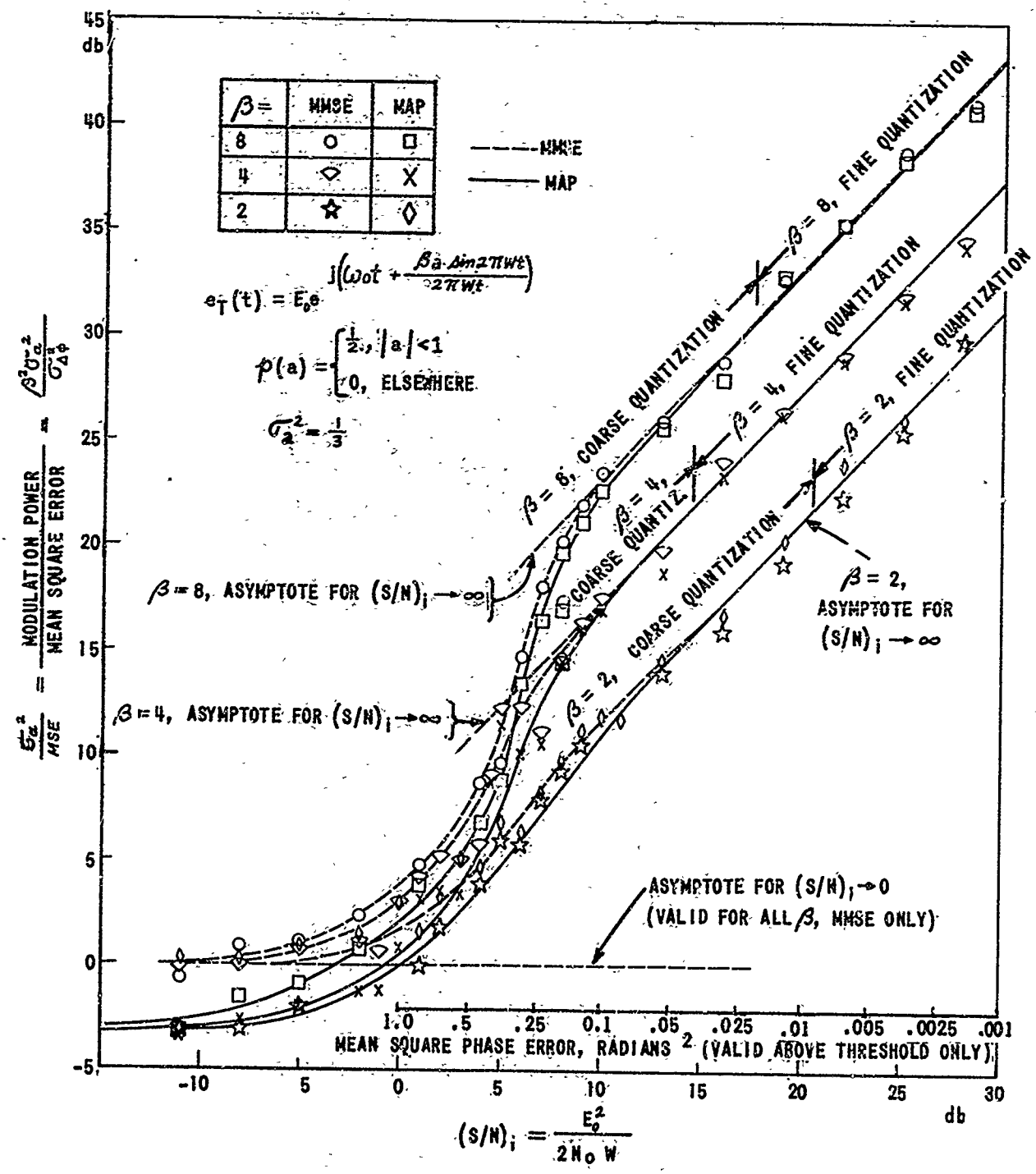


Figure IV-1 PLOTS OF $\frac{\sigma_a^2}{MSE}$ vs $(S/N)_i$

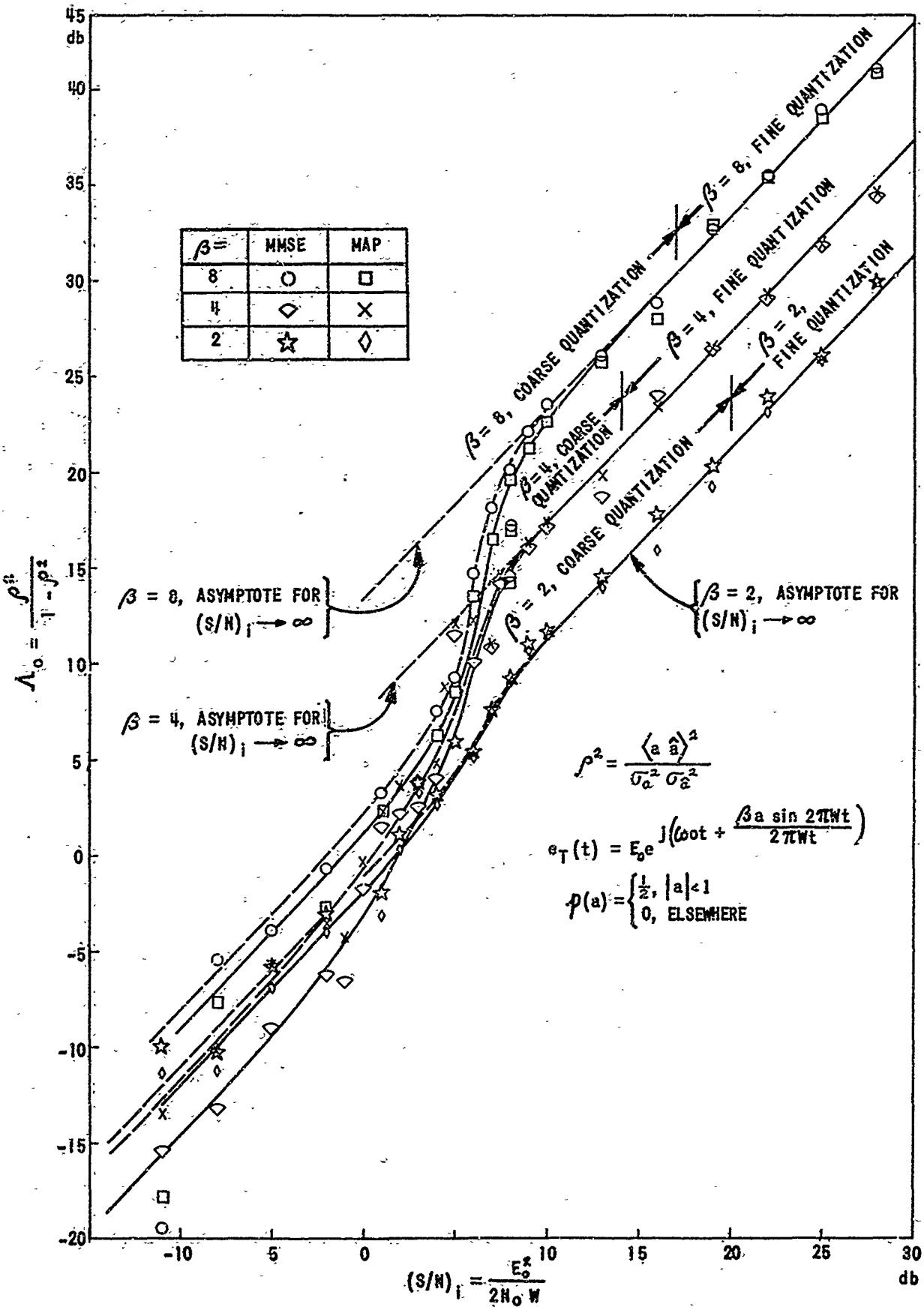


Figure IV-2 PLOTS OF $\Lambda_o = \frac{\rho^2}{1-\rho^2}$ vs $(S/N)_i$

is seen to be excellent. This result is important because the experiment was performed in a theoretically optimum manner, while the analytical expression was obtained by the use of several simplifying assumptions. Since a phase detector followed by an optimum (Wiener) filter has the same asymptotic behavior as derived above,[2] it follows that improvement of performance through the use of more sophisticated demodulation techniques cannot be obtained in the large $(S/N)_i$ region where the asymptotic results apply. The most important goal from a practical point of view is to approach the asymptotic performance at lower $(S/N)_i$, i.e., to extend the threshold as far as possible toward lower $(S/N)_i$. Threshold is usually defined as that value of input signal-to-noise ratio below which the output signal-to-noise ratio is no longer linearly related to the input signal-to-noise ratio. Therefore, in this program threshold occurs at a value of $(S/N)_i$ below which $\frac{\sigma_a^2}{MSE}$ begins to deviate significantly below the value predicted by the linear asymptotic formula While the asymptotic formula gives no indication of where threshold will occur, one may speculate that the threshold occurs when the mean-square filtered phase error reaches a critical value. From the discussion above the mean-square filtered phase error expressed in decibels is linearly related to the ordinate of Figure IV-1. This relationship can be expressed as

$$\left(\sigma_{\Delta\phi}^2\right)_{db} = \left[\beta^2\,\sigma_a^2\right]_{db} - \left[\frac{\sigma_a^2}{MSE}\right]_{db} . \qquad (5)$$

The experimental data suggests that threshold occurs at $(S/N)_i \approx$ 7 db, which corresponds to a value of $\sigma_{\Delta\phi}^2 = .2$ radians^2 as predicted by the linear theory. The threshold effect is seen to become much more pronounced, i.e., the departure from linearity becomes more rapid as the modulation index β is increased. There is indication that the threshold value of $(S/N)_i$ increases ($\sigma_{\Delta\phi}^2$ decreases) somewhat as β increases; this behavior was predicted on theoretical grounds in Chapter I.

There appear to be two contributary components which cause the threshold. The first is the nonlinear increase in output noise as the input signal-to-noise ratio is decreased. This phenomenon occurs when the assumption $|n_1|, |n_2| \ll E_o$ is no longer tenable. The second component is caused by "ambiguities", that is, errors in $\hat{\phi}$ by integer multiples of 2π. With the modulation restricted to the range [-1, +1] ambiguities can be encountered only when $\beta > \pi$ with either maximum a posteriori (MAP) or

minimum mean-square error (MMSE) estimation. A detailed discussion of the MAP and MMSE estimation procedure is presented below. For the present it suffices to note that with either procedure the estimate of the modulation cannot fall outside the range [-1, +1].* With PM and an observation of the received signal at one instant of time only, it is possible to determine the phase only modulo 2π. In the absence of noise the resulting ambiguities can be resolved by examining the time history of the received signal. When a small amount of noise is added, the resulting phase estimate is perturbed, but no difficulty in resolving the ambiguities is encountered. The estimates of phase deteriorate at first in direct proportion to the noise power and then more rapidly; this is the first indication of threshold. As the noise is increased still more the ambiguities can no longer be resolved and large errors begin to occur. Because of the noise these large errors are not exact multiples of 2π. We have adopted the convention of calling an error, the magnitude of which exceeds π radians, a click. Figure IV-3 is a plot of the click probability with MAP estimation, versus intrinsic signal-to-noise ratio with $\beta = 8$. Comparison with Figure IV-1, IV-2, shows that clicks set in at an input signal-to-noise ration in the vicinity of threshold. The extremely rapid increase in click probability causes the rapid deterioration of the output estimates encountered below threshold when large values of modulation index are used. As the input signal-to-noise ratio continues to decrease the continuum noise and the clicks due to ambiguities merge. At moderate signal-to-noise ratios, just below threshold, it is, however, easy to differentiate the clicks from the continuum noise. The set of histograms, Figures IV-4, IV-5, give a good illustration of how the output error distribution changes as a function of input signal-to-noise ratio.**

---

*In general the MMSE estimate $\hat{a}$ can take on a value where the prior probability density $p(a)$ is zero. However, this is not possible when the prior distribution is non-zero over a connected set, as is the case with a uniform distribution.

**Only the results of MAP estimation are shown in Figures IV-3, 4, 5. The reason for this is that the computational effort required for the MAP case is considerably less than for the corresponding MMSE case because the estimates always coincide with the quantized values of $a(o)$ in the MAP case.

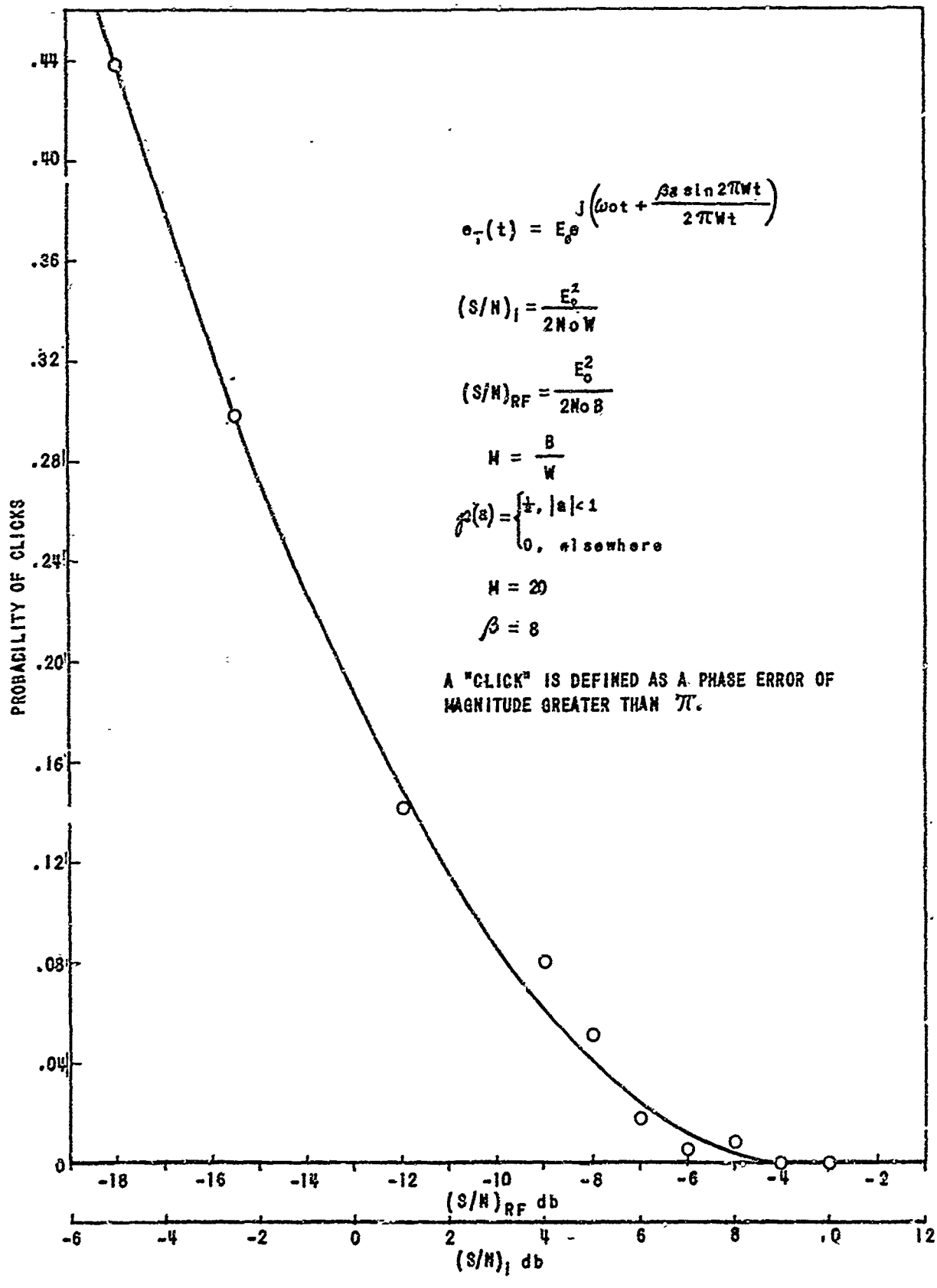
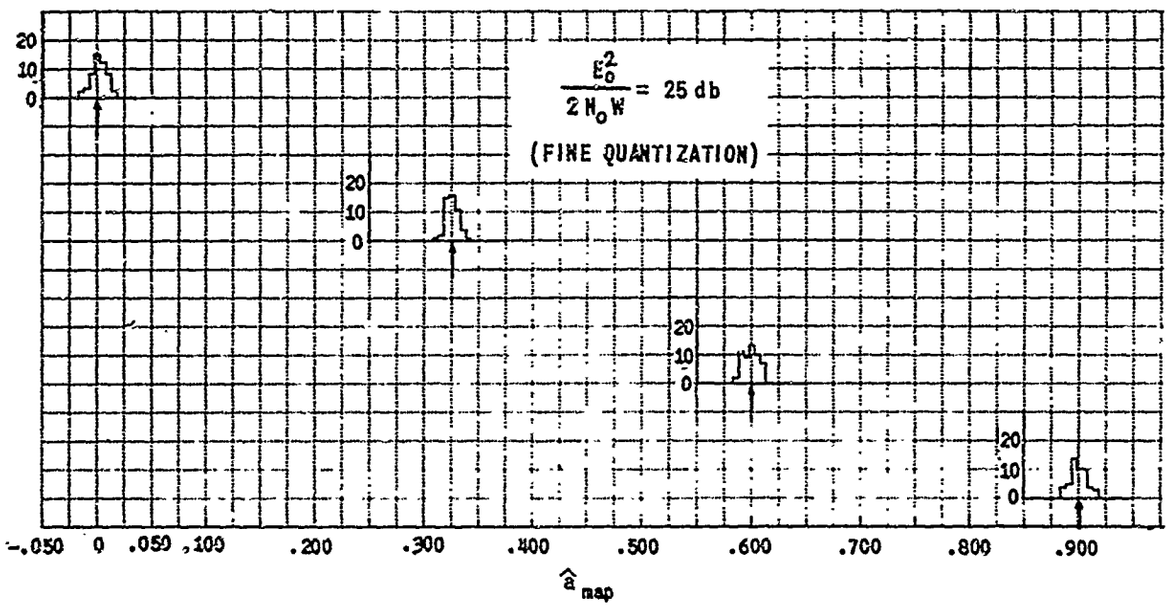
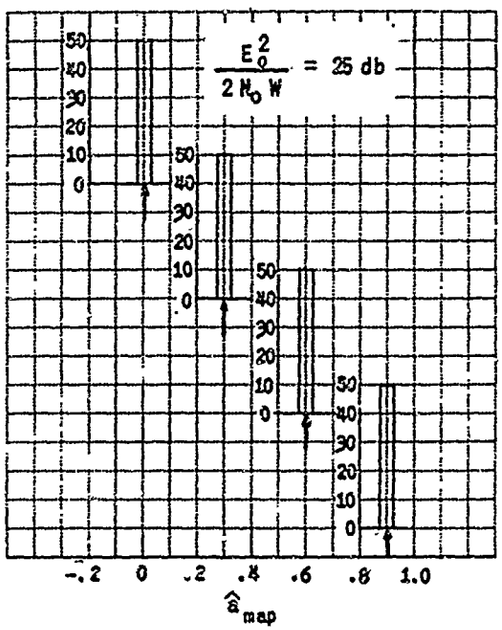


Figure IV-3 PROBABILITY OF CLICKS vs (S/N) db, $\beta = 8$ (MAP ESTIMATION)





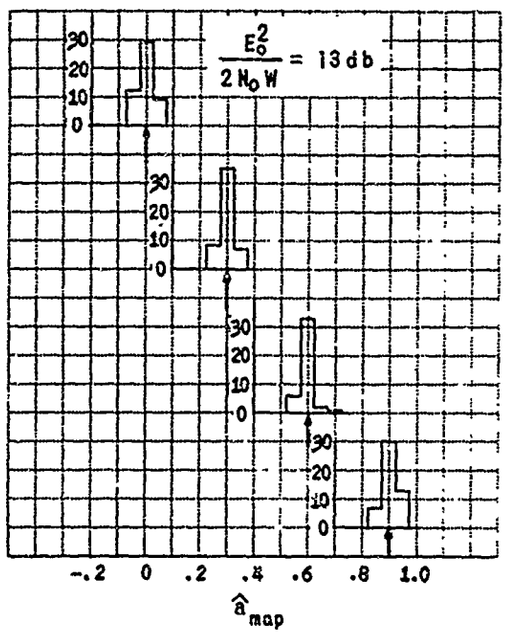


Figure V-4 HISTOGRAMS OF 50 TRIALS. $\beta = 8$, M = 20. MAP ESTIMATION. (ARROWS INDICATE AMPLITUDE OF TRANSMITTED SIGNALS)

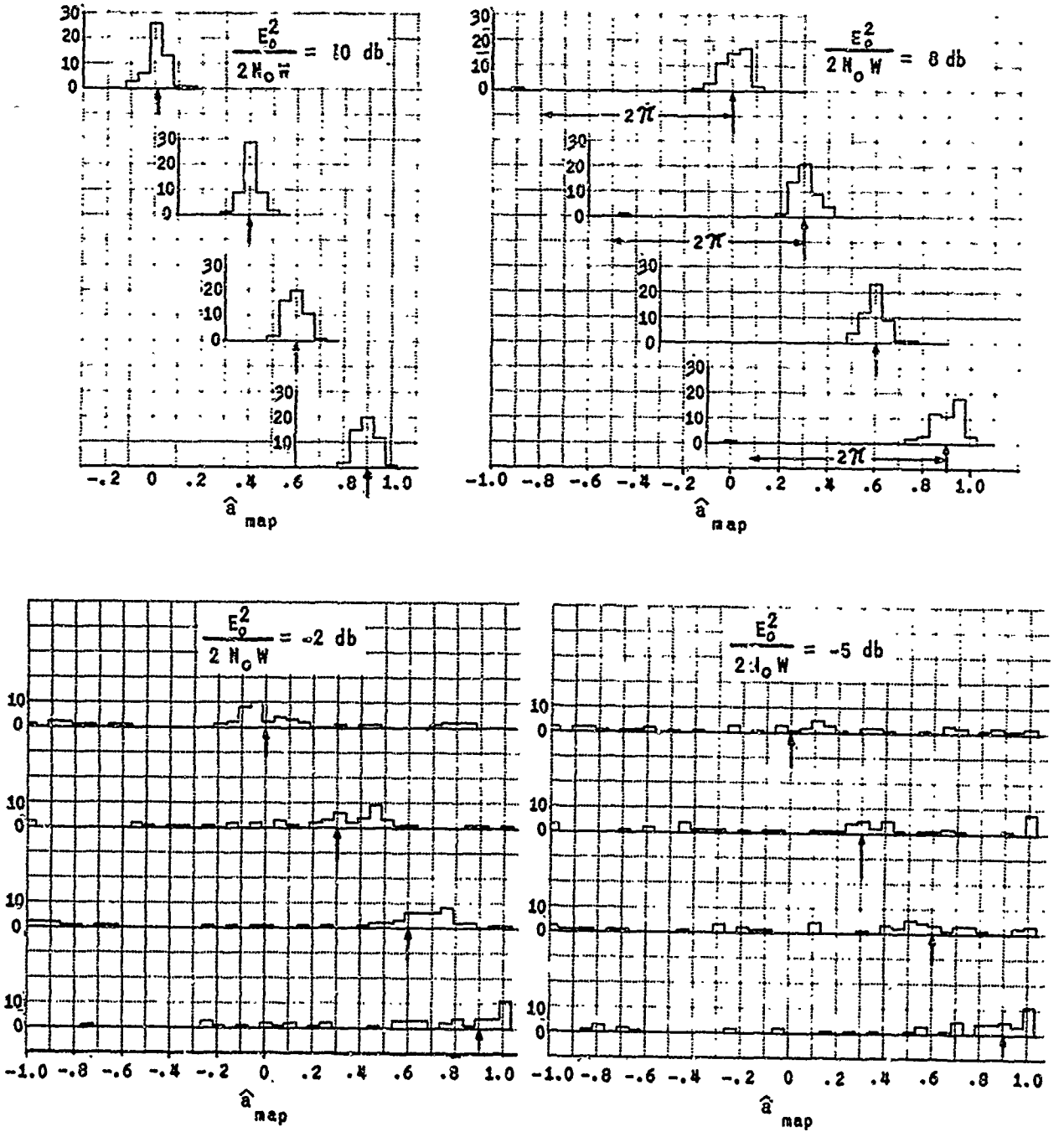


Figure IV-5 HISTOGRAMS OF 50 TRIALS. $\beta = 8$, $H = 20$. MAP ESTIMATION. (ARROWS INDICATE AMPLITUDE OF TRANSMITTED SIGNALS)

Theoretical Background and Experimental Program

The performance of the MAP and MMSE estimators is evaluated either in terms of the mean-square error or the parameter $\Lambda_0 = \frac{\rho^2}{1-\rho^2}$ These performance measures can be evaluated empirically by a Monte Carlo technique, provided that the estimate $\hat{a}$ can be determined for all noise perturbed signals. A sufficiently large number N of pairs $(a, \hat{a})$ are generated in accordance with the appropriate signal and noise statistics and the sample mean-square error, the correlation coefficient and Λ_0 are computed as follows:

$$MSE = \frac{1}{N} \sum_{j=1}^{N} (a_j - \hat{a}_j)^2,$$

$$\rho^2 = \frac{\left(\sum_{j=1}^{N} a_j \hat{a}_j\right)^2}{\sum_{j=1}^{N} a_j^2 \sum_{j=1}^{N} \hat{a}_j^2} \quad \text{(It is assumed that } \langle a \rangle = 0, \langle \hat{a} \rangle = 0 \text{)} \tag{6}$$

We assume that the modulation a(t) is strictly bandlimited to the band $0 - W$, Hz, a(t) is then recoverable from samples spaced $\frac{1}{2W}$ seconds apart by means of the interpolation formula

$$a(t) = \sum_{i} a\left(\frac{i}{2W}\right) \frac{\sin \pi (2Wt - i)}{\pi (2Wt - i)} \tag{7}$$

Therefore, only the values of $a\left(\frac{i}{2W}\right)$ need be estimated. In principle the entire set of $\left\{\hat{a}\left(\frac{i}{2W}\right)\right\}$ is to be chosen, based on the observed received waveform and the prior statistics of $\left\{a\left(\frac{i}{2W}\right)\right\}$ and the additive noise. The complexity of the resulting simulation grows exponentially with the number of samples which must be considered. Clearly, if the value of all of the samples except one were known, the estimate of the remaining sample could only be improved and the complexity of the simulation would be greatly reduced. We obtained upper bounds on the attainable performance by setting $a\left(\frac{i}{2W}\right) = 0$ for all $i \neq 0$. In an actual system where the values of $a\left(\frac{i}{2W}\right), i \neq 0$ are not known, the errors obtained in estimating $a(0)$ would therefore be greater than those observed here.

The problem was thus reduced to obtaining an optimum estimate based on observation of the received signal $e_R(t) = e_T(t) + n(t)$.

The transmitted signal is assumed to be phase modulated in accordance with $e_T[a(t)] = E_o \cos[\omega_o t + \beta_o a(t)]$. Thus the RF phase, in the absence of modulation, is also known to the receiver. (This phase can be ascertained by averaging over a sufficiently long time.)

In the simulation it is a practical necessity to work with low frequency equivalents of the actual RF signals. For this purpose it is convenient to use the in-phase and quadrature components (using $\cos \omega_o t$ as the reference) of the RF signals. The rate at which these components can change is determined by the RF bandwidth. Limiting the RF bandwidth to B Hz permits the in-phase and quadrature components to be reconstructed from samples spaced $1/B$ seconds apart. Increasing the RF bandwidth thus requires more samples to be taken (and consequently increases the cost of computation) but can only lead to improved performance. However, the improvement can be expected to level off rapidly once a certain point has been reached. A method of choosing the RF bandwidth as a function of modulation index will be described subsequently.

With an RF bandwidth of B Hz, the sampled values of the in-phase and quadrature components of received signals can be generated as follows (this is the same technique as used in the analyses of Chapters II and III):

$$e_R(t) = e_1 \cos \omega_o t - e_2(t) \sin \omega_o t$$

where

$$\begin{aligned} e_1(t) &= E_o \cos \beta a(t) + n_1(t) \\ e_2(t) &= E_o \sin \beta a(t) + n_2(t) \\ \text{let } \Delta &= 1/B \\ e_1(K\Delta) &= E_o \cos \beta a(K\Delta) + n_1(K\Delta) \\ e_2(K\Delta) &= E_o \sin \beta a(K\Delta) + n_2(K\Delta) \\ n_{1K} &= n_1(K\Delta) \\ n_{2K} &= n_2(K\Delta) \\ a_K &= a(K\Delta) \end{aligned} \qquad (8)$$

Assuming that the noise is white and Gaussian over the RF bandwidth, with intensity N_0 watts/Hz, n_{1K}, n_{2K} are independent zero mean Gaussian variables with variance $N_0 B$ *. Since the modulation function $a(o) \frac{\sin 2\pi Wt}{2\pi Wt}$ approaches zero asymptotically for large $|t|$ and all finite values of $a(o)$ it is necessary to limit the observation interval; however, the performance is hardly degraded when the higher order sidelobes are eliminated. Examination of the $\frac{\sin x}{x}$ function[5] indicated that an observation interval which includes the main lobe and the first two sidelobes, on each side of the main lobe, should be satisfactory. This conclusion was experimentally verified by observation of the dependence of the MSE which was obtained as the length of the observation interval was varied. The results obtained are plotted in Figure IV-6 for intrinsic input signal-to-noise ratios $(S/N)_i$ of +6 and +22 db corresponding to operation just below and well above threshold. From the figure it is apparent that extending the observation interval beyond the first two sidelobes does not improve the performance appreciably.

A uniform a priori distribution of $a(o)$ over the range [-1, +1] was used. Therefore, the phase deviation at $t = o$ was uniformly distributed over $[-\beta, +\beta]$ radians. In order to carry through the digital evaluation it is necessary to quantize this range. The magnitude of the quantum steps is determined by the more stringent of the following requirements. The quantization steps must result in errors which are negligible compared to the errors due to the thermal noise and the quantization steps of the phase angle must be sufficiently small so as to present a good approximation of the sine and cosine functions. At low $(S/N)_i$ where the second requirement is more stringent, 20 equal quantum steps were used with $\beta \leq 4$ and 40 steps with $\beta = 8$. This corresponds to a maximum phase increment of .4 rad $\approx$ 22.8°. At large $(S/N)_i$ where the quantization noise with 20 (or 40) steps would be comparable to or greater than the errors due to the thermal noise, the number of quantum steps was increased tenfold to 200 or 400 equal steps. These two cases are differentiated on Figures IV-1, 2, 4, 6

*The in-phase and quadrature noise components are independent provided the noise spectrum is even about ω_o.

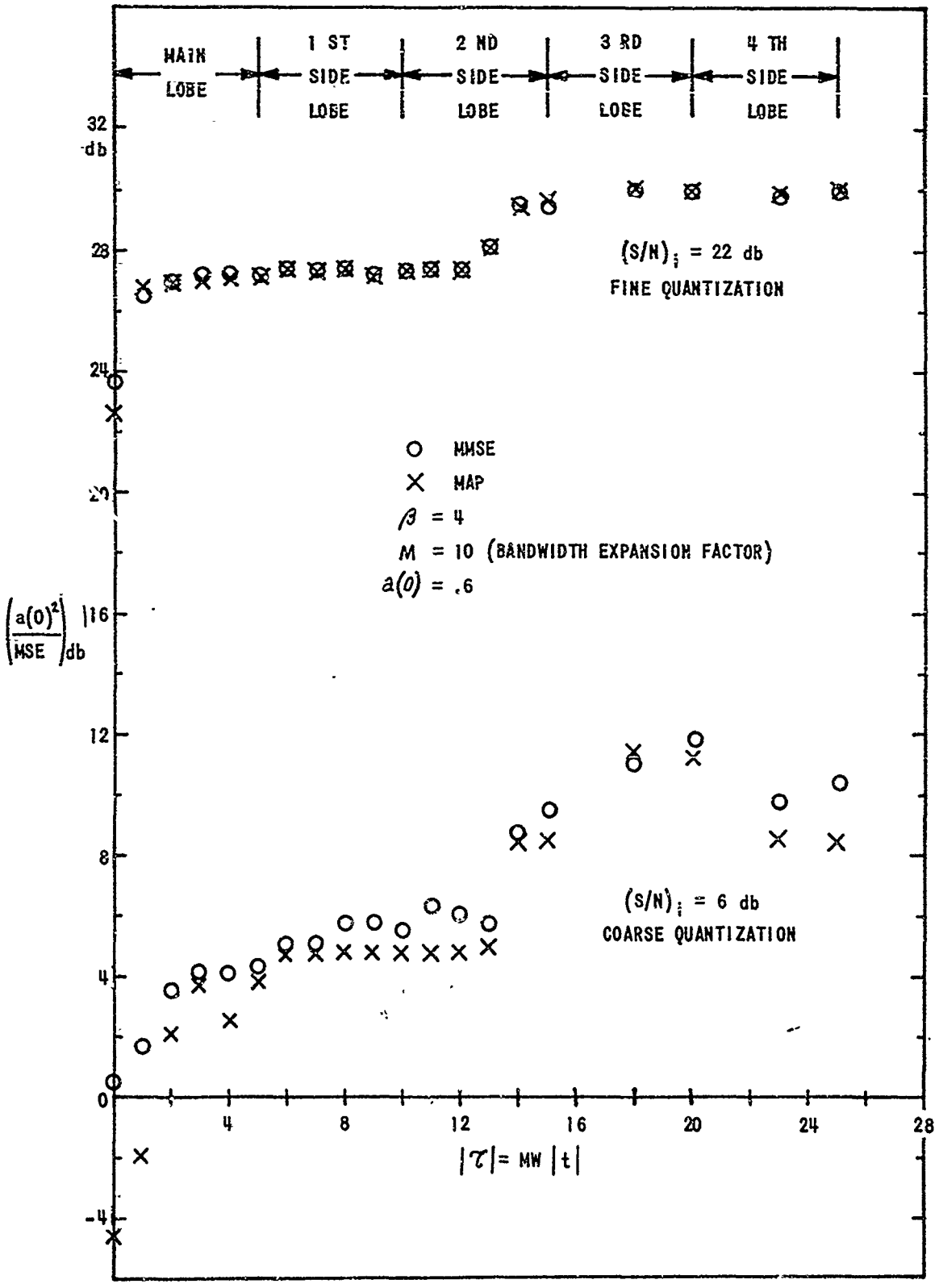


Figure IV-6 MSE vs OBSERVATION INTERVAL $2|\tau|$

by the terms "coarse" and "fine" quantization. Figure IV-1 for $(S/N)_i = 25$ db, illustrates why it is necessary to go to a finer quantization with large $(S/N)_i$. In this case, when the coarse quantization, 21 sample values of $a(0)$, hence 20 quantum steps, was used, the MAP estimate of $a(0)$ always coincided with the correct value of $a(0)$, the mean-square error would therefore have been zero. However, when the number of sample values of $a(0)$ was increased to 401 the error assumed a distribution centered about the true value and the mean-square error observed had a value in agreement with the asymptotic theory.

Let us briefly review the criteria by which the maximum likelihood, the maximum a posteriori probability and the minimum mean-square error estimators determine their estimates.*

1. Maximum Likelihood Estimation

For any received signal $e_R(t)$ the maximum likelihood estimator designates that modulation $\hat{a}_1(t)$ which maximizes the conditional probability of obtaining the observed received signal. This is the probability that the noise will be such that when it is added to the transmitted signal $e_T(a,t)$ which results from modulation with a(t), the observed received signal will be obtained. $\hat{a}_1(t)$ is therefore that function which maximizes $p(e_R/a)$ as a function of $a(t)$ It is to be noted that knowledge of the statistics of the modulating signal, i.e., its prior probability density $p(a)$ is not required in order to obtain the maximum likelihood estimate. All that is required is knowledge of the modulation process and the noise statistics.

2. Maximum a Posteriori Probability, MAP, Estimation

For any received signal $e_R(t)$ the maximum a posteriori probability estimator designates that modulation, $\hat{a}_2(t)$ which maximizes its a posteriori probability density. $\hat{a}_2(t)$ is therefore that function which maximizes $p(a/e_R)$. The relationship between the maximum likelihood

*The nomenclature is not consistent throughout the literature. In particular, the terms maximum likelihood and maximum a posteriori probability estimator are frequently used synonymously, viz. Ref. [1, 2].

estimator and the maximum a posteriori probability estimator can be developed with the aid of Bayes theorem

$$p(a, e_R) = p(a/e_R)\,p(e_R) = p(e_R/a)\,p(a)$$

$$p(a/e_R) = \frac{p(e_R/a)\,p(a)}{p(e_R)}$$

For a given received signal $e_R(t)$

$$p(a/e_R) = K\,p(e_R/a)\,p(a)$$

from which it is seen that the conditional probability density of $e_R(t)$ given a(t) (which is maximized by the maximum likelihood estimator) must be multiplied by the a priori probability density of the modulation in order to determine the maximum a posteriori probability estimate. Note that the maximum likelihood and the maximum a posteriori probability estimators perform identically when $p(a)$ is a constant.

3. Minimum Mean Square Error, MMSE, Estimation

For every received signal $e_R(t)$ the minimum mean-square error estimator designates that modulation $\hat{a}_3(t)$ which results in the minimum conditional mean-square error. $\hat{a}_3(t)$ is therefore that function which minimizes $\langle(\hat{a}_3 - a)^2 / e_R\rangle$

It is easily shown that $\hat{a}_3(t) = \langle a(t)/e_R(t)\rangle$ = conditional expectation of a(t) given $e_R(t)$,

viz. $$\langle(\hat{a}_3 - a)^2 / e_R\rangle = \hat{a}_3^2 - 2\hat{a}_3\langle a/e_R\rangle + \langle a^2/e_R\rangle$$

$$\frac{\partial\langle(\hat{a}_3 - a)^2/e_R\rangle}{\partial\hat{a}_3} = 2\hat{a}_3 - 2\langle a/e_R\rangle = 0$$

$$\hat{a}_3 = \langle a/e_R\rangle \qquad (9)$$

For all of these estimators the major problem is to evaluate the likelihood function $p(e_R/a)$ as a function of a(t).

For the case of additive noise the probability density $p(e_R/a)$ that a particular received signal $e_R(t)$ is the result of adding $n(t)$ to a

particular transmitted signal $e_T(t)$ is equal to the probability that the noise $n(t) = e_R(t) - e_T(t)$ For white Gaussian bandlimited noise

$$p(n(\cdot)) = K \exp\left[-\frac{1}{N_0}\int n^2(\tau)\,d\tau\right] = K \exp\left[-\frac{1}{N_0}\int \left\{e_r(\tau) - e_T(\tau)\right\}^2 d\tau\right]$$

$$= K \exp\left[-\frac{1}{N_0}\int \left\{e_r(\tau) - e_T\left[a(\tau),\tau\right]\right\}^2 d\tau\right] = p(e_r \mid a) \qquad (10)$$

where $e_T\left[a(\tau),\tau\right]$ is the RF signal corresponding to the modulation $a(\tau)$. In principle the likelihood function can be determined empirically by means of the above formula. One would select a sufficiently dense set of $a(\tau)$ and evaluate the likelihood function at each of these points, each point corresponding to a particular function $a(\tau)$ in a high dimensional signal space.

The difficulty with this approach is that with typical parameter angle modulation, even for relatively simple statistics of $p(a)$, the number of candidate functions $a(\tau)$ required for evaluation of a single likelihood function is untractably large. By restricting the modulation function to be of the form $a(0)\,\frac{\sin 2\pi W t}{2\pi W t}$ the problem is reduced from a vector estimation to a scalar estimation problem; instead of estimating the function a(t) only its value, $a(0)$, at time $t = 0$ need be estimated.

It has already been mentioned that both theory and experiment indicate that an observation interval covering the main lobe and two sidelobes (on each side) of the $\frac{\sin 2\pi W t}{2\pi W t}$ function should suffice for the current investigation. The main lobe of this function has a width of $1/W$ and each of the sidelobes has a width $1/2W$ so that this observation interval has a duration $T = \frac{6}{2W}$. With an RF bandwidth B the RF signal can be reproduced from samples of the in-phase and quadrature components spaced at intervals $\Delta = 1/B$ seconds apart. Thus, placing one sample at $t = 0$ there are $TB + 1 = 3\,\frac{B}{W} + 1 = 3M + 1 = L$ samples of the in-phase and 3 M+1 samples of the quadrature component required to adequately describe the RF waveform over this interval.* The RF waveform can then be

*This statement is not exact because a signal cannot be strictly limited in both time and frequency. For large bandwidth time products the error becomes negligible.

represented as a point in a $2L$ dimensional Cartesian coordinate system. In this waveform space the noise has a spherically symmetric Gaussian distribution with zero mean. Changing the amplitude of $a(0)$ causes the signal point to trace a convoluted line through this space, but this signal graph is nowhere tightly packed into the space. In the estimation procedure the continuous distribution $p(a)$ is approximated with a discrete distribution. With the adjacent values of $a(0)$ chosen such that the increments in phase angle, $\Delta\phi = \beta\Delta\, a(0)$ are less than .4 radians,* in order to insure a sufficiently dense sampling of the RF signal space along the signal graph. The coordinates of the corresponding set of points in the signal space are then computed and the likelihood function $p(e_R/a)$ is evaluated at these points in accordance with

$$p(e_R/a) = K \exp\left[-\frac{1}{N_0 B}\sum_K \left\{(e_{1K} - E_0 \cos \beta a_K)^2 + (e_{2K} - E_0 \sin \beta a_K)^2\right\}\right]$$

$$= K \exp\left[-\frac{1}{N_0 B}\, d^2(a)\right] \qquad (11)$$

Where $d(a)$ is the distance between the transmitted waveform $e_T[a,t]$ resulting from the modulation a(t), and the received waveform $e_R(t)$ The likelihood function is zero at any point in the signal space which does not lie on the signal graph. The important concept is that every waveform limited to the RF band of width B Hz centered on frequency $\frac{\omega_0}{2\pi}$ and duration T can be represented in this signal space, but that possible transmitted signals are constrained to lie on the signal graph. By restricting our attention to this graph the magnitude of the problem has therefore been greatly reduced.

With the likelihood function defined along the signal graph the estimates derived in accordance with the three criteria described earlier, can be determined.

---

*As mentioned previously, the number of signal points was increased at high (S/N)$_i$ in order to keep the effects of the quantization negligible in comparison with the errors due to noise.

It is interesting to observe that the maximum likelihood estimate corresponds simply to the smallest value of $d^2(a)$ and hence does not require evaluation of $p(e_R/a)$ or knowledge of $N_0 B$ This comment also applies to MAP estimation with $p(a)$ constant as we have assumed here. The MMSE estimate, however, requires the evaluation of $exp\left[-\frac{1}{N_0 B} d^2(a)\right]$ and therefore knowledge of $N_0 B$. The bandwidth B used in this simulation was chosen by a conservative application of the so-called Carson's rule

$$B = 2\left(W + \frac{1}{2\pi}\left(\frac{d\phi}{dt}\right)_{max}\right)$$

$$\text{with } \phi = \beta a(o)\,\frac{\sin 2\pi W t}{2\pi W t}$$

$$\frac{1}{2\pi}\frac{d\phi}{dt} = \beta a(o)\, 2\pi W \frac{d}{dx}\left(\frac{\sin x}{x}\right) .$$

The function $\frac{\sin x}{x}$ and its derivatives have been tabulated,[5] its first derivative attains a maximum value of .43618 at x $\cong$ - 119.5°.

Since use of a larger bandwidth than required can result only in improved estimates but at the expense of increased computation, we have for the purposes of the application of Carson's rule assumed $\left(\frac{d\phi}{dt}\right)_{max} = 2\pi W\beta$ (i.e., we have replaced 'a' by its maximum value of unity and used 1.0 instead of .43618). So that $B = 2W(\beta+1)$ and the bandwidth expansion factor $M = 2(\beta+1)$. For convenience in the actual computations $M = 10$ was used for all $\beta \leq 4$, and $M = 20$ for $\beta = 8$.

Conclusion

The performance of a phase modulated system was investigated by means of a digital computer. For the case of a simplified modulation process the computer evaluation of the performance of estimators (receivers) optimized in accordance with the MMSE and MAP criteria was evaluated. This method readily permitted investigation of the behavior of these estimators in and below the threshold region, where analytic investigations are very difficult.

In applying the results obtained one should particularly bear in mind that the curves are upper bounds to the performance obtainable with non-realtime estimators and that the modulation was uniformly distributed in amplitude. It would be very desirable to extend the investigation to Gaussianly distributed modulation processes, where a comparison with analytically derived results (Chapters II, III) and experimental results (Chapter V) could be more easily made. Although no difficulty is apparent in modifying the computer program to handle this case, there was insufficient time available to undertake these further investigations.

APPENDIX IV-A
COMPUTER PROGRAM

```
C     ****************************************************************
C
C     DECK NAME. DEMOA
C
C     MACHINE. IBM 7044
C
C     SOURCE LANGUAGE. FORTRAN IV
C
C     PURPOSE
C        MONTE CARLO EVALUATION OF THE PERFORMANCE OF PM ANGLE MODULATION
C        SYSTEMS
C
C     DESCRIPTION OF INPUT PARAMETERS
C         NSNR      - NUMBER OF SIGNAL-TO-NOISE RATIOS
C         NATS      - NUMBER OF VALUES OF AO USED
C         NRNS      - NUMBER OF STATISTICAL TRIALS
C         NTTS      - NUMBER OF CANDIDATE TRANSMITTED SIGNALS
C         NTMS      - NUMBER OF TIME SAMPLES ON EACH SIDE OF CENTRAL
C                     SAMPLE.TOTAL NUMBER OF TIME SAMPLES FOR THE IN PHASE
C                     COMPONENT IS 2(NTMS)+1.TOTAL NUMBER OF TIME SAMPLES
C                     FOR THE QUADRATURE COMPONENT IS 2(NTMS)+1
C         NSTART    - TWELVE DIGIT OCTAL INTEGER. INTIAL STATE OF RANDOM
C                     NUMBER GENERATOR
C         BETA      - MODULATION INDEX
C         BEF       - BANDWIDTH EXPANSION FACTOR
C         SNRIDB    - SIGNAL-TO-NOISE RATIO IN DECIBELS
C         JZ        - INDEX INDICATING POSITION OF TRANSMITTED SIGNAL
C                     IN ARRAY OF CANDIDATE SIGNALS
C         AO        - AMPLITUDE OF TRANSMITTED MODULATION
C         PAO       - PRIOR PROBABILITY OF AO
C         AZ        - AMPLITUDE OF CANDIDATE SIGNAL
C         PA        - PRIOR PROBABILITY OF AZ
C
C     DESCRIPTION OF SYMBOLS USED IN PRINTOUT
C         S/N       - SIGNAL-TO-NOISE RATIO
C         (S/N)DB   - SIGNAL-TO-NOISE RATIO IN DECIBELS
C         A(O)      - AMPLITUDE OF TRANSMITTED MODULATION
C         RNS       - NUMBER OF TRIAL
C         MLE       - MAXIMUM A POSTERIORI ESTIMATE
C         MMSE      - MINIMUM MEAN SQUARE ERROR ESTIMATE
C         MVO       - SAMPLE MEAN ESTIMATE
C         MVO2      - SAMPLE MEAN SQUARE ESTIMATE
C         SIGO2     - SAMPLE VARIANCE OF ESTIMATES
C         MSE       - MEAN SQUARE ERROR
C         PROB      - PROBABILITY OF CORRECT IDENTIFICATION OF A(O) IN MAP
C                     MODE
C         AVE.MSE   - WEIGHTED AVERAGE MSE
C
C     REMARKS
C        PROGRAM PRODUCES PUNCHED CARD OUTPUT.FIRST CARD IDENTIFIES OUTPUT.
C        A PROGRAM STOP OCCURS WHEN PROB=1.0
C
C     SUBPROGRAMS REQUIRED
C        DVCHK     - CAL SYSTEM LIBRARY SUBPROGRAM CU 0032.TESTS FOR A
C                    ZERO DIVISOR
C        GRNORG(N) -CAL SYSTEM LIBRARY SUBPROGRAM CU 0024.SETS INITIAL
C                   STATE OF GAUSSIAN RANDOM NUMBER GENERATOR TO THE
C                   VALUE N
C        GRNTP(X,N)-CAL SYSTEM LIBRARY SUBPROGRAM CU 0024.GENERATES N
C                   PSEUDO RANDOM GAUSSIAN DEVIATES(FLOATING POINT,ZERO
C                   MEAN,UNIT VARIANCE) IN LOCATIONS X(1),...,X(N)
C
C     ****************************************************************
```

```
C
C      SSDEMOA
       DIMENSION AO(4),PAO(4),Q1(2),Q2(2),      Q3(2),Q4(2),Q5(2),Q6(2),
      1Q7(2),Q8(2),Q9(2),Q10(2),Q11(2),JZ(4),SEST(2),SEST2(2),
      2SINCX(30),ER(61 ),EN(61 ),A(2),AZ(41,4),Z(41),ET(61,41,4),
      3RHO2(2),Q12(2),SNRIDB(25),PA(41,4)
       COMPLEX ET,ER,EN,IMAG,DIFF
       DATA SS/6HSSDATA/
       CALL DVDCHK(1)
       PUNCH 302,SS
  302  FORMAT(72X,A6)
  500  READ(5,1) NSNR,NATS,NRNS,NTTS,NTMS,NSTART
    1  FORMAT (5I10,D12)
       READ(5,12) BETA,BEF
   12  FORMAT (8E10.0)
       READ (5,12) (SNRIDB(I),I=1,NSNR)
       READ(5,1) (JZ(I),I=1,NATS)
       READ (5,12) (AO(I),I=1,NATS)
       READ (5,12) (PAO(I),I=1,NATS)
       DO 400 J=1,NATS
       READ (5,72)  (AZ(I,J),I=1,NTTS)
       READ (5,72)  (PA(I,J),I=1,NTTS)
   72  FORMAT (16F5.0)
  400  CONTINUE
       WRITE(6,2)
    2  FORMAT (53H1PERFORMANCE OF ANGLE MODULATED SIGNALS - J.LAWTON/SS/)
       WRITE (6,3) NSNR,NATS,NRNS,NTTS,NTMS,NSTART
    3  FORMAT (6HONSNR=,I3,2X,5HNATS= ,I3,2X,5HNRNS=,I3,2X,5HNTTS=,I3,2X,
      15HNTMS=,I3,2X,7HNSTART=,O12)
       WRITE (6,13) BETA,BEF
   13  FORMAT (6HOBETA=,F7.2,2X,4HBEF=,F7.2)
       WRITE (6,14) (SNRIDB(I),I=1,NSNR)
   14  FORMAT (7HOSNRIDB/1X,15F8.1)
       WRITE(6,75) (JZ(I),I=1,NATS)
   75  FORMAT (3HOJZ/1X,10I4)
       WRITE (6,70) (AO(I),I=1,NATS)
   70  FORMAT (4HOAO=,6E20.8)
       WRITE (6,344) (PAO(I),I=1,NATS)
  344  FORMAT (5HOPAO=,6E20.8)
       DO 300 J=1,NATS
       WRITE (6,73) (AZ(I,J),I=1,NTTS)
   73  FORMAT (3HOAZ/(1X,16F8.3))
       WRITE (6,74)  (PA(I,J),I=1,NTTS)
   74  FORMAT (3HOPA/(1X,16F8.4))
  300  CONTINUE
       PI=3.1415926536
       FNRNS=FLOAT(NRNS)
       IMAG=CMPLX(0.0,1.0)
       NIQS=2*NTMS+1
       NGRN=2*NIQS
       FNTTS=FLOAT(NTTS)
       DO 10 I=1,NTMS
       R=(2.0*FLOAT(I))/BEF
       IF (R.EQ.AINT(R)) GO TO 11
       X=PI*R
       SINCX(I)=SIN(X)/X
       GO TO 10
   11  SINCX(I)=0.0
   10  CONTINUE
```

```
      DO 23  NTS=1,NATS
      DO 20 J=1,NTTS
      NADD=0
      DO 21 NREP=1,2
      DO 22 I=1,NTMS
      K=I+NADD
      PHI=SETA*AZ(J,NTS)*SINCX(I)
      ET(K,J,NTS)=COS(PHI)-IMAG*SIN(PHI)
   22 CONTINUE
      NADD=NTMS
   21 CONTINUE
      PHI=BETA*AZ(J,NTS)
      ET(NIQS ,J,NTS)=COS(PHI)-IMAG*SIN(PHI)
   20 CONTINUE
   23 CONTINUE
      IF (NSTART.NE.0)   CALL GRNORG(NSTART)
      DO 30 NR=1,NSNR
      DO 62 NE=1,2
      Q5(NE)=0.0
      Q6(NE)=0.0
      Q8(NE)=0.0
   62 CONTINUE
      SNR=10.0**(SNRIDB(NR)/10.0)
      BNZ=0.5/SNR
      SIGMA=SQRT(BNZ)
      DO 31 NTS=1,NATS
      PROB=0.0
      JATS=JZ(NTS)
      AZA=AO(NTS)
      PAZ=PAO(NTS)
      QZ=PAZ*AZA
      WRITE (6,4) SNR,SNRIDB(NR)
    4 FORMAT (5H1S/N=,F8.3,3X,8HS/N(DB)=,E11.3)
      AZA=AO(NTS)
      WRITE  (6,5) AZA
    5 FORMAT  (6HOA(O)=,F8.3)
      WRITE (6,6)
    6 FORMAT (1HO,9X,3HRNS,5X,3HMLE,19X,4HMMSE)
      DO 40 NE=1,2
      SEST(NE)=0.0
      SEST2(NE)=0.0
   40 CONTINUE
      DO 32 NT=1,NRNS
      CALL GRNTP(EN,NGRN)
      DO 33 I=1,NIQS
      ER(I)=ET(I,JATS,NTS)+SIGMA*EN(I)
   33 CONTINUE
      A1MAX=0.0
      SA3DK=0.0
      SPA2DK=0.0
      DO 34 J=1,NTTS
      DIS=0.0
      DO 35 I=1,NIQS
      DIFF=ER(I)-ET(I,J,NTS)
      DIS=DIS+REAL(DIFF)**2+AIMAG(DIFF)**2
   35 CONTINUE
      Z(J)=DIS/BNZ
   34 CONTINUE
      ZMIN=Z(1)
      DO 341 J=2,NTTS
      IF (Z(J).LT.ZMIN) ZMIN=Z(J)
  341 CONTINUE
```

```
      DO 342 J=1,NTTS
      PA1DK=EXP(ZMIN-Z(J))
      IF (PA1DK.GT.A1MAX) GO TO 36
      GO TO 37
   36 A1MAX=PA1DK
      JA1=J
   37 PA2DK=PA1DK*PA(J,NTS)
      SPA2DK=SPA2DK+PA2DK
      SA3DK=SA3DK+PA2DK*AZ(J,NTS)
  342 CONTINUE
      A(1)=AZ(JA1,NTS)
      IF (A(1).EQ.AZA)      PROB=PROB+1.0
      A(2)=SA3DK/SPA2DK
      WRITE (6,7) NT,(A(NE),NE=1,2)
    7 FORMAT (10X,I2,F11.3,11X,F11.3/)
      DO 50 NE=1,2
      SEST(NE)=SEST(NE)+A(NE)
      SEST2(NE)=SEST2(NE)+A(NE)**2
   50 CONTINUE
   32 CONTINUE
      PROB=PROB/FNRNS
      DO 60 NE=1,2
      Q1(NE)=SEST(NE)/FNRNS
      Q2(NE)=SEST2(NE)/FNRNS
      Q3(NE)=Q2(NE)-Q1(NE)**2
      Q4(NE)=(AZA-Q1(NE))**2+Q3(NE)
      Q5(NE)=Q5(NE)+QZ*Q1(NE)
      Q6(NE)=Q6(NE)+PAZ*Q2(NE)
      Q8(NE)=Q8(NE)+PAZ*Q4(NE)
   60 CONTINUE
      WRITE (6,92) (Q1(NE),NE=1,2)
   92 FORMAT (10X,5HMVO =,E12.3,E22.3)
      WRITE (6,93) (Q2(NE),NE=1,2)
   93 FORMAT(9X,6HMVO2 =,E12.3,E22.3)
      WRITE (6,94) (Q3(NE),NE=1,2)
   94 FORMAT (8X,7HSIGO2 =,E12.3,E22.3)
      WRITE (6,95) (Q4(NE),NE=1,2)
   95 FORMAT (10X,5HMSE =,E12.3,E22.3)
      WRITE (6,200) PROB
  200 FORMAT (9X,6HPROB =,E12.3)
      IF(PROB.EQ.1.0) STOP
   31 CONTINUE
      DO 90 NE=1,2
      RHO2(NE)= Q5(NE)**2/(.3375*Q6(NE))
      Q7(NE)=RHO2(NE)/(1.0-RHO2(NE))
      Q12(NE)=10.0*ALOG10(RHO2(NE))
      Q10(NE)=10.0*ALOG10(Q7(NE))
      Q9(NE)=0.3375/Q8(NE)
      Q11(NE)=10.0*ALOG10(Q9(NE))
   90 CONTINUE
      WRITE (6,104) (RHO2(NE),NE=1,2)
  104 FORMAT (9X,6HRHO2 =,E12.3,E22.3)
      WRITE (6,105) (Q12(NE),NE=1,2)
  105 FORMAT (11X,4HDB =,E12.3,E22.3)
      WRITE (6,96) (Q7(NE),NE=1,2)
   96 FORMAT (2X,13HRHO2/1-RHO2 =,E12.3,E22.3)
      WRITE (6,105) (Q10(NE),NE=1,2)
      WRITE (6,97) (Q8(NE),NE=1,2)
   97 FORMAT (5X,10HAVE. MSE =,E12.3,E22.3)
      WRITE(6,98) (Q9(NE),NE=1,2)
   98 FORMAT (2X,13H.3375/AVMSE =,E12.3,E22.3)
      WRITE (6,105) (Q11(NE),NE=1,2)
      PUNCH 301, SNRIDB(NR),Q10(1),Q10(2),Q11(1),Q11(2)
  301 FORMAT (F8.1,4E18.3)
   30 CONTINUE
      GO TO 500
      END
```

APPENDIX IV-B
SUMMARY OF COMPUTED RESULTS

β = 2, BEF = 10, NTMS = 15

$(S/N)_i$ db*	$\Lambda_o = \frac{\rho^2}{1-\rho^2}$ MAP db	MMSE db	$\frac{\sigma_a^2}{MSE}$ MAP db	MMSE db
-11.0	-0.113E 02	-0.946E 01	-0.318E 01	0.337E 00
-8.0	-0.112E 02	-0.102E 02	-0.384E 01	0.603E-01
-5.0	-0.692E 01	-0.575E 01	-0.341E 01	0.827E 00
-2.0	-0.400E 01	-0.300E 01	-0.845E 00	0.144E 01
-1.0	-0.523E 01	-0.197E 01	-0.235E-01	0.142E 01
2.0	[illegible] 00	0.105E 01	0.195E 01	0.372E 01
3.0	[illegible] 01	0.361E 01	0.391E 01	0.516E 01
4.0	0.269E 01	0.325E 01	0.381E 01	0.470E 01
5.0	0.587E 01	0.695E 01	0.597E 01	0.603E 01
6.0	0.513E 01	0.546E 01	0.574E 01	0.619E 01
7.0	0.747E 01	0.763E 01	0.733E 01	0.833E 01
8.0	0.902E 01	0.923E 01	0.924E 01	0.965E 01
9.0	0.107E 02	0.110E 02	0.105E 02	0.113E 02
10.0	0.113E 02	0.117E 02	0.115E 02	0.119E 02
13.0	0.140E 02	0.145E 02	0.148E 02	0.147E 02
16.0	0.159E 02	0.167E 02	0.160E 02	0.168E 02
19.0	0.192E 02	0.203E 02	0.192E 02	0.236E 02
23.0	0.224E 02	0.218E 02	0.229E 02	0.234E 02
25.0	0.254E 02	0.267E 02	0.259E 02	0.267E 02
22.0	0.231E 02	0.219E 02	0.231E 02	0.239E 02
55.0	0.258E 02	0.260E 02	0.259E 02	0.260E 02
28.0	0.298E 02	0.299E 02	0.293E 02	0.299E 02

β = 4, BEF = 10, NTMS = 15

$(S/N)_i$ db*	Λ_o MAP db	MMSE db	$\frac{\sigma_a^2}{MSE}$ MAP db	MMSE db
-11.5	-0.153E 02	-0.135E 02	-0.372E 01	-0.107E 00
-8.0	-0.132E 02	-0.103E 02	-0.275E 01	0.173E-01
-5.0	-0.897E 01	-0.613E 01	-0.191E 01	0.685E 00
-2.0	-0.623E 01	-0.335E 01	-0.123E 01	0.114E 01
-1.0	-0.657E 01	-0.446E 01	-0.121E 01	0.426E 01
0.0	-0.103E 01	0.189E 00	0.957E 00	0.298E 01
1.0	0.145E 01	0.229E 01	0.317E 01	0.427E 01
2.0	0.757E 01	0.168E 01	0.354E 01	0.517E 01
3.0	0.457E 01	0.400E 01	0.347E 01	0.814E 01
4.0	0.400E 01	0.475E 01	0.496E 01	0.578E 01
4.5	0.668E 01	0.671E 01	0.845E 01	0.912E 01
5.0	0.115E 02	0.120E 02	0.114E 02	0.123E 02
6.0	0.999E 01	0.122E 02	0.132E 02	0.124E 02
7.0	0.107E 02	0.109E 02	0.137E 02	0.113E 02
8.0	0.142E 02	0.145E 02	0.143E 02	0.146E 02
9.0	0.160E 02	0.162E 02	0.151E 02	0.163E 02
10.0	0.171E 02	0.174E 02	0.171E 02	0.175E 02
13.0	0.187E 02	0.198E 02	0.197E 02	0.198E 02
16.0	0.222E 02	0.231E 02	0.229E 02	0.231E 02
19.0	0.257E 02	0.263E 02	0.257E 02	0.263E 02
16.0	0.234E 02	0.240E 02	0.234E 02	0.240E 02
19.0	0.264E 02	0.265E 02	0.262E 02	0.264E 02
22.0	0.290E 02	0.292E 02	0.293E 02	0.291E 02
25.0	0.318E 02	0.317E 02	0.314E 02	0.319E 02
28.0	0.343E 02	0.346E 02	0.347E 02	0.346E 02

β = 8, BEF = 20, NTMS = 30

$(S/N)_i$ db*	Λ_o MAP db	MMSE db	$\frac{\sigma_a^2}{MSE}$ MAP db	MMSE db
-11.0	-0.178E 02	-0.145E 02	-0.159E 01	-0.545E 00
-8.0	-0.767E 01	-0.940E 01	-0.156E 01	0.931E 00
-5.0	-0.584E 01	-0.196E 01	-0.971E 00	0.108E 01
-2.0	-0.257E 01	-0.125E 00	0.731E 00	0.240E 01
1.0	0.242E 01	0.347E 01	0.379E 01	0.473E 01
4.0	0.626E 01	0.590E 01	0.649E 01	0.409E 01
5.0	0.857E 01	0.924E 01	0.896E 01	0.965E 01
6.0	0.135E 02	0.147E 02	0.139E 02	0.144E 02
7.0	0.165E 02	0.181E 02	0.165E 02	0.181E 02
8.0	0.169E 02	0.172E 02	0.173E 02	0.173E 02
8.0	0.142E 02	0.146E 02	0.145E 02	0.147E 02
8.0	0.196E 02	0.202E 02	0.196E 02	0.207E 02
9.0	0.213E 02	0.221E 02	0.211E 02	0.230E 02
10.0	0.227E 02	0.235E 02	0.227E 02	0.235E 02
13.0	0.256E 02	0.261E 02	0.256E 02	0.260E 02
16.0	0.280E 02	0.288E 02	0.283E 02	0.288E 02
19.0	0.326E 02	0.327E 02	0.324E 02	0.327E 02
22.0	0.354E 02	0.355E 02	0.354E 02	0.354E 02
25.0	0.385E 02	0.369E 02	0.394E 02	0.388E 02
28.0	0.408E 02	0.411E 02	0.406E 02	0.411E 02
31.0	0.436E 02	0.441E 02	0.436E 02	0.441E 02

COARSE STRUCTURE — β = 4, BEF = 10

NTMS	$(S/N)_i$ db	$\frac{\sigma_a^2}{MSE}$ MAP db	MMSE db
25	6.0	0.845E 01	0.104E 02
20	6.0	0.112E 02	0.118E 02
15	6.0	0.849E 01	0.945E 01
10	6.0	0.469E 01	0.549E 01
7	6.0	0.467E 01	0.512E 01
5	6.0	0.379E 01	0.431E 01
4	6.0	0.254E 01	0.404E 01
3	6.0	0.371E 01	0.417E 01
2	6.0	0.206E 01	0.348E 01
1	6.0	-0.190E 01	0.173E 01
0	6.0	-0.453E 01	0.466E 00

FINE STRUCTURE

NTMS	$(S/N)_i$ db	MAP db	MMSE db
25	25.0	0.299E 02	0.300E 02
20	25.0	0.300E 02	0.299E 02
15	25.0	0.296E 02	0.294E 02
10	25.0	0.273E 02	0.273E 02
7	25.0	0.274E 02	0.273E 02
5	25.0	0.272E 02	0.272E 02
4	25.0	0.270E 02	0.272E 02
3	25.0	0.269E 02	0.272E 02
2	25.0	0.269E 02	0.270E 02
1	25.0	0.265E 02	0.265E 02
0	25.0	0.228E 02	0.237E 02

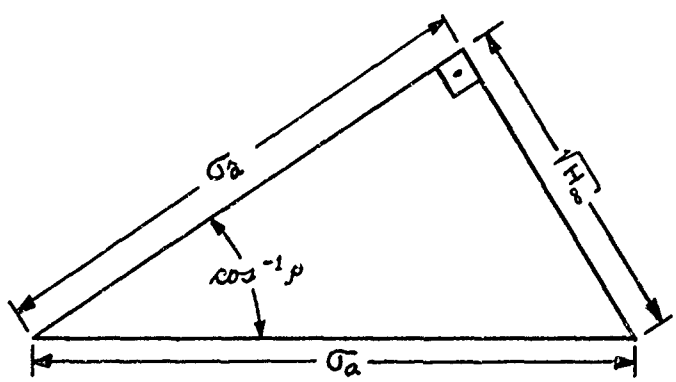
* NOTE: RESULTS ARE TABULATED AS A FUNCTION OF $(S/N)_i = \frac{E_i^2}{2N_oW}$ COMPUTATIONS PERFORMED IN TERMS OF SNR $= \frac{E_s^2}{2N_oB} = \frac{W}{B}(S/N)_i$, $\mu = \frac{B}{W}$ = BEF = BANDWIDTH EXPANSION FACTOR.

APPENDIX IV-C

LINEAR ANALYSIS OF MMSE ESTIMATION OF PM AND FM MODULATED SIGNALS

Phase Modulation

Comparison of Equation (22) Chapter III with Equation (4) Chapter IV shows that the low signal-to-noise asymptote and the high signal-to-noise asymptote coincide in the limit as β approaches zero. (For $(S/N)_i >> 1$, $\Lambda_o \sim \frac{\sigma_a^2}{MSE}$). One naturally inquires whether the attainable performance is for all signal-to-noise ratios given by these asymptotes. That this is indeed so is demonstrated below. For any β the mean-square phase deviation due to modulation $\langle \theta^2 \rangle = \beta^2 \sigma_a^2$. For very small β the mean-square phase deviation is very small and the approximations $\sin\theta \sim \tan\theta \sim \theta$ and $\cos\theta \sim 1$ can be used. In the limit as $\beta \rightarrow 0$, phase modulation thus becomes a linear modulation in quadrature to the unmodulated carrier. Let the modulation "a" have a flat spectrum from 0-W Hz and the RF noise be white and Gaussian with intensity N_0 Watts/Hz. The quadrature component of the received RF signal is then given by $\beta E_0 a + n_2$ where n_2 is a white Gaussian random process of intensity $2N_0$ Watts/Hz. (one-sided spectrum). The minimum mean-square error in estimating a(t) is obtainable by use of a quadrature detector, with gain $1/\beta E_0$ to recover "a", followed by an optimum (infinite delay Wiener) filter to separate the noise. Let $\hat{a}$ represented the MMSE estimate of "a" and H_∞ the resulting mean-square error. $H_\infty = \langle (a - \hat{a})^2 \rangle$. The minimization of the mean-square error requires $\langle \hat{a}(a - \hat{a}) \rangle = 0$. The relationship between the correlation coefficient ρ and the normalized mean-square error H_∞, shown in the sketch, is derived below



$$\left\langle a(a-\hat{a})\right\rangle = 0 \tag{C-1}$$

$$\left\langle \hat{a}\, a\right\rangle = \left\langle \hat{a}^2\right\rangle = \sigma_{\hat{a}}^2 \tag{C-2}$$

$$\begin{aligned} H_\infty = \left\langle (a-\hat{a})^2\right\rangle &= \left\langle a^2\right\rangle - 2\left\langle a\,\hat{a}\right\rangle + \left\langle \hat{a}^2\right\rangle \\ &= \left\langle a^2\right\rangle - \left\langle \hat{a}^2\right\rangle = \sigma_a^2 - \sigma_{\hat{a}}^2 \end{aligned} \tag{C-3}$$

$$\begin{aligned} \rho^2 = \frac{\left\langle a\,\hat{a}\right\rangle^2}{\sigma_a^2\,\sigma_{\hat{a}}^2} = \frac{\sigma_{\hat{a}}^2}{\sigma_a^2} &= 1 - \frac{\sigma_a^2 - \sigma_{\hat{a}}^2}{\sigma_a^2} = 1 - \frac{H_\infty}{\sigma_a^2} \\ &= 1 - h \end{aligned} \tag{C-4}$$

$$\Lambda_0 = \frac{\rho^2}{1-\rho^2} = \frac{1-h}{h} = \frac{1}{h} - 1 \tag{C-5}$$

For the optimum filter $H_\infty = \int_{-\infty}^{+\infty} \frac{S(\omega)\,N(\omega)}{S(\omega)+N(\omega)}\,\frac{d\omega}{2\pi}$

$$= \int_{-2\pi W}^{2\pi W} \frac{\frac{1}{2W}\,\frac{N_0}{(E_0\beta)^2}}{\frac{1}{2W} + \frac{N_0}{(E_0\beta)^2}}\,\frac{d\omega}{2\pi}$$

$$= \frac{\frac{N_0}{(E_0\beta)^2}}{\frac{1}{2W} + \frac{N_0}{(E_0\beta)^2}} = h \quad \text{since } \sigma_a^2 = 1 \tag{C-6}$$

so that

$$\Lambda_o = \frac{\rho^2}{1-\rho^2} = \frac{\frac{1}{2W} + \frac{N_o}{(E_o\beta)^2}}{\frac{N_o}{(E_o\beta)^2}} - 1 = \frac{\beta^2 E_o^2}{2N_o W} = \beta^2 (S/N)_i \qquad \text{(C-7)}$$

It is important to realize that the above expression is correct asymptotically as $\beta \to 0$, at all values of $(S/N)_i$. For $(S/N)_i \gg 1$, $\Lambda_o = \frac{\rho^2}{1-\rho^2} \sim \frac{\sigma_a^2}{MSE} = \frac{1}{h}$. Comparison with Equation (117) Chapter II or Equation (4) Chapter IV shows that this result, Equation (C-7), is also asymptotically valid for all β as $(S/N)_i \longrightarrow \infty$. Comparison with Equation (22) Chapter III shows agreement for $\beta \to 0$ and $(S/N)_i \to 0$.

Frequency Modulation

The performance of an FM system can also be determined by linear analysis provided that σ_θ^2, the variance of the phase deviation due to modulation is sufficiently small.* In the frequency modulation case σ_θ^2 depends on both the modulation index β and the spectrum of $S_a(\omega)$ of the modulation, because with FM the phase deviation is inversely proportional to the modulating frequency. (We assume as before that the modulation has zero mean, $\langle a \rangle = 0$, and unity variance,
$\sigma_a^2 = \int_{-\infty}^{+\infty} S_a(\omega) \frac{d\omega}{2\pi} = 1$.)

Consider a bandpass spectrum

$$S_a(\omega) = \begin{cases} \frac{\pi}{\omega_2 - \omega_1}, & |\omega_1| < |\omega| < |\omega_2| \\ 0, & \text{elsewhere} \end{cases} \qquad \text{(C-8)}$$

$$\theta = \beta \int a(t)\, dt \qquad \text{(C-9)}$$

$$S_\theta(\omega) = \frac{\beta^2 S_a(\omega)}{\omega^2} = \begin{cases} \frac{\beta^2 \pi}{\omega^2(\omega_2 - \omega_1)}, & |\omega_1| < |\omega| < |\omega_2| \\ 0, & \text{elsewhere} \end{cases} \qquad \text{(C-10)}$$

*For $(S/N)_i$ sufficiently large so that the in-phase component of the noise can be neglected, a linear analysis of FM can be carried through for all σ_θ^2 in the same manner as for PM.

$$\langle \theta^2 \rangle = \int_{-\infty}^{+\infty} S_\theta(\omega) \frac{d\omega}{2\pi} = \frac{\beta^2}{\omega_2 - \omega_1} \int_{\omega_1}^{\omega_2} \frac{1}{\omega^2} d\omega$$

$$= \frac{\beta^2}{\omega_2 - \omega_1} \left(\frac{1}{\omega_1} - \frac{1}{\omega_2} \right) = \frac{\beta^2}{\omega_2 \omega_1} = \sigma_\theta^2 \qquad \text{(C-11)}$$

For this spectrum the linear analysis applies provided $\frac{\beta^2}{\omega_1 \omega_2} << 1$.

The received phase angle

$$\phi \sim \frac{n_2}{E_0} + \theta = \frac{n_2 + \beta E_0 \int a(t)\, dt}{E_0} \qquad \text{(C-12)}$$

$$\frac{1}{\beta} \phi' \sim \frac{n_2'}{\beta E_0} + a(t) \qquad \text{(C-13)}$$

where a prime denotes $\frac{d(\ )}{dt}$.

The mean-square error in the estimation of a(t) at the output of a Wiener filter is then

$$H_\infty = \int_{-\infty}^{+\infty} \frac{\frac{1}{\beta^2 E_0^2} S_{n_2'}(\omega)\ S_a(\omega)}{\frac{1}{\beta^2 E_0^2} S_{n_2'}(\omega) + S_a(\omega)} \frac{d\omega}{2\pi} \qquad \text{(C-14)}$$

With $S_a(\omega)$ given by Equation (C-8) and $S_{n_2}(\omega) = N_0$ (two-sided spectrum)

$$H_\infty = \int_{\omega_1}^{\omega_2} \frac{\frac{2N_0}{(\beta E_0)^2}\omega^2 \frac{2\pi}{\omega_2-\omega_1}}{\frac{2N_0}{(\beta E_0)^2}\omega^2 + \frac{2\pi}{\omega_2-\omega_1}} \cdot \frac{d\omega}{2\pi}$$

$$= \frac{1}{\omega_2-\omega_1}\int_{\omega_1}^{\omega_2} \frac{\omega^2}{\omega^2 + \frac{2\pi(\beta E_0)^2}{2N_0(\omega_2-\omega_1)}}\, d\omega$$

$$= 1 - \frac{1}{\omega_2-\omega_1}\int_{\omega_1}^{\omega_2} \frac{\frac{2\pi(\beta E_0)^2}{2N_0(\omega_2-\omega_1)}}{\omega^2 + \frac{2\pi(\beta E_0)^2}{2N_0(\omega_2-\omega_1)}}\, d\omega$$

$$= 1 - \frac{\sqrt{\frac{2\pi(\beta E_0)^2}{2N_0(\omega_2-\omega_1)}}}{(\omega_2-\omega_1)}\left[\tan^{-1}\left(\frac{\omega_2}{\sqrt{\frac{2\pi(\beta E_0)^2}{2N_0(\omega_2-\omega_1)}}}\right) - \tan^{-1}\left(\frac{\omega_1}{\sqrt{\frac{2\pi(\beta E_0)^2}{2N_0(\omega_2-\omega_1)}}}\right)\right]$$

$$= 1 - \sqrt{\frac{\beta^2(S/N)_i}{\omega_1\omega_2\left(\rho - 2 + \frac{1}{\rho}\right)}}\left[\tan^{-1}\left(\frac{1}{\sqrt{\frac{\beta^2}{\omega_1\omega_2}\frac{(S/N)_i}{\rho}}}\right) - \tan^{-1}\left(\frac{1}{\sqrt{\frac{\beta^2}{\omega_1\omega_2}(S/N)_i\rho}}\right)\right]$$

$$= 1 - \sqrt{\frac{\beta^2(S/N)_i}{\omega_1\omega_2\left(\rho-2+\frac{1}{\rho}\right)}}\left[\alpha - \gamma\right], \text{ where } \quad \rho = \frac{\omega_2}{\omega_1}, \; (S/N)_i = \frac{E_0^2}{2N_0\left(\frac{\omega_2-\omega_1}{2\pi}\right)}$$

$$\left[\alpha - \gamma\right] = \tan^{-1}\left[\frac{\tan\alpha - \tan\gamma}{1 + \tan\alpha\tan\gamma}\right]$$

$$H_\infty = 1 - \sqrt{\frac{\beta^2(S/N)_i\,\rho}{\omega_1\omega_2(\rho-1)^2}}\tan^{-1}\left(\sqrt{\frac{\beta^2(S/N)_i(\rho-1)^2}{\omega_1\omega_2\,\rho\left(1+\frac{\beta^2}{\omega_1\omega_2}(S/N)_i\right)^2}}\right)$$

$$= 1 - \sqrt{\sigma_\theta^2(S/N)_i\frac{\rho}{(\rho-1)^2}}\tan^{-1}\left(\sqrt{\frac{\sigma_\theta^2(S/N)_i(\rho-1)^2}{\left(1+\sigma_\theta^2(S/N)_i\right)^2\rho}}\right) \qquad \text{(C-15)}$$

Substituting Equation (C-15) into Equation (C-5) plots of Λ_o vs $(S/N)_i$ were prepared over the range $-40 \leq (S/N)_i \leq 50$ db and for all combinations of the parameters $\sigma_\theta^2 = \frac{\beta^2}{\omega_1\omega_2} = .01, .1, .3, 1$ and $\rho = \frac{\omega_2}{\omega_1} = 1.01, 10, 100$.

The resulting curves are reproduced as Figure III-3.

Asymptotic solutions for the large and the small $(S/N)_i$ cases can be obtained by limiting operations on the integrand of Equation (C-14), or by series expansions of the arc-tangent function.

We thus obtain

$$\Lambda_0 \sim \frac{3\beta^2}{\omega_2 \omega_1 (P+1+1/P)} (S/N)_i = \frac{3\sigma_\theta^2}{P+1+1/P} (S/N)_i \,, \quad (S/N)_i \gg 1 \tag{C-16}$$

$$\Lambda_0 \sim \frac{\beta^2}{\omega_2 \omega_1} (S/N)_i = \sigma_\theta^2 (S/N)_i \,, \quad (S/N)_i \ll 1 \tag{C-17}$$

in agreement with Equation (27) of Chapter III.

APPENDIX IV-D

SENSITIVITY OF $\frac{\sigma_a^2}{MSE}$ AND Λ_0 ORDINATES OF FIGURE IV-1, IV-2, TO EXPERIMENTAL ERRORS

The sample mean-square error obtained experimentally will be an inexact estimate of the mean-square error. In plotting $\Lambda_{0\,db}$ vs $(S/N)_{i\,db}$ the sensitivity of $\Lambda_{0\,db}$ to errors in the value of the mean-square error goes to infinity as $(S/N)_i$ goes to zero, i.e., as the mean-square error approaches σ_a^2. Thus,

$$\Lambda_0 = \frac{\rho^2}{1-\rho^2} = \frac{1-\frac{MSE}{\sigma_a^2}}{1-\left(1-\frac{MSE}{\sigma_a^2}\right)} = \frac{1-h}{h} \tag{D-1}$$

where $h = \frac{MSE}{\sigma_a^2}$ represents the minimum mean-square error normalized with respect to σ_a as discussed in Appendix IV-C.

$$\Lambda_{0\,db} = 10\left[\log(1-h) - \log h\right] \tag{D-2}$$

$$\frac{d\Lambda_{0\,db}}{dh} = -10\left[\frac{1}{1-h} + \frac{1}{h}\right] = \frac{-10}{h(1-h)} \tag{D-3}$$

$$h_{db} = 10 \log h\,, \quad \frac{dh_{db}}{db} = \frac{10}{h} \tag{D-4}$$

$$\frac{d\Lambda_{0\,db}}{dh_{db}} = \frac{d\Lambda_{0\,db}}{dh}\,\frac{1}{\frac{dh_{db}}{dh}} = -\frac{1}{1-h} \to -\infty \text{ as } h \to 1, \left(h_{db} \to 0\right) \tag{D-5}$$

On the other hand, plots of $\left(\frac{\sigma_a^2}{MSE}\right)_{db}$ vs $(S/N)_{i\,db}$ are not subject to this difficulty. Thus,

$$\frac{d\left(\frac{\sigma_a^2}{MSE}\right)_{db}}{dh_{db}} = \left(-10\,h\right)\frac{h}{10} = -h^2 \to -1 \text{ as } h \to 1, \left(h_{db} \to 0\right) \tag{D-6}$$

Because of the above dependence, plots of the experimentally obtained $\left(\frac{\sigma_a^2}{MSE}\right)_{db}$ vs $(S/N)_{i\,db}$. Figure IV-1 show less scatter in the low $(S/N)_i$ region than the corresponding plots of $\Lambda_{0\,db}$ vs $(S/N)_{i\,db}$. Figure IV-2.

REFERENCES

[1] Youla, D.C., "The Use of the Method of Maximum Likelihood in Estimating Continuous-Modulated Intelligence Which Has Been Corrupted by Noise", IRE Transactions PGIT-3, pp. 90-105, March 1954.

[2] Lawton, John G., "Investigation of Analog and Digital Communication Systems" (Phase 3 Report), RADC-TDR-63-147, ASTIA Document No. AD-407343, May 1963.

[3] Van Trees, Harry L., "Analog Communication Over Randomly-Time-Varying Channels", IEEE Transactions on Information Theory, Vol. 12, No. 1, pp. 51-63, January 1966.

[4] Viterbi, Andrew J., "Principles of Coherent Communication", Chapter 5, pp. 123-147, McGraw-Hill Book Company, 1966.

[5] "Tables of the Function $\sin\phi/\phi$ and its First Eleven Derivations", by the Staff of the Computation Laboratory (Harvard University), Harvard University Press, 1949.

V. COMPUTER SIMULATION OF FM DEMODULATION; WITH EMPHASIS ON PLL OPERATION IN THE THRESHOLD REGION

1. Introduction

The relationship of phase-locked loop (PLL) FM demodulator design to the theory of maximum a posteriori (MAP) estimation of FM signals is described and the threshold characteristic of a PLL system based on our previous MAP analysis[1]* is evaluated by means of a hybrid computer simulation. The results are compared against the well established performance of limiter-discriminator receivers, other PLL evaluations, and the theoretical work on the limit to threshold extension as developed by Akima.[2]

It is the intent of this simulation to specify and evaluate a physically realizable demodulator that will give, with some reasonable criteria, the best possible estimate of the modulation in a system in which the carrier is phase modulated by some function of the information to be transmitted. In the generalized angle modulation system shown in Figure V-1, the linear operator would be a constant multiplier for a phase modulation system, whereas an FM system would require that this operator be an integrator.

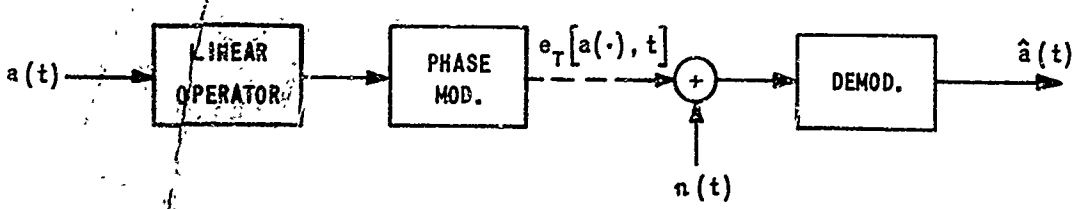


Figure V-1

While much of what follows is extendable to any linear operator, the work will be specialized to the design and performance evaluation of FM demodulators.

---

*Early work in this field by Youla (Reference [3]) introduced the term "maximum likelihood", which, although somewhat of a misnomer, was used in our previous work (Reference [1]). As discussed in greater detail in Chapter IV, the term maximum a posterior (MAP) is more appropriate and will henceforth be used in this Chapter.

If a(t) is the modulating signal (i.e., the information to be transmitted) the instantaneous angular frequency of the transmitted signal may be written

$$\omega(t) = \omega_0 + \beta a(t) \tag{1}$$

and the resulting phase is

$$\phi(t) = \int_0^t \omega(\tau)\, d\tau = \omega_0 t + \beta \int_0^t a(\tau)\, d\tau \tag{2}$$

Thus, the FM signal is of the form

$$e_T(t) = E_0 \cos\left[\omega_0 t + \beta \int_0^t a(\tau)\, d\tau\right] \tag{3}$$

The most obvious approach to extracting the modulation signal, $a(\tau)$ from the received signal is to perform essentially the inverse of the modulation process. That is, the demodulator would generate an output proportional to the instantaneous frequency (or rate of change of phase) of the received signal. This is precisely what the limiter-discriminator or zero-crossing demodulator is designed to do. The performance of the discriminator type demodulator has been quite well established from both theoretical analyses and empirical results. At sufficiently high carrier-to-noise power ratios, the signal-to-noise ratio at the demodulator output is linearly proportional to the input carrier-to-noise ratio. However, a marked threshold is observed when the carrier-to-total noise power at the input to the discriminator is approximately 10-12 db; below threshold the output signal-to-noise ratio decreases rapidly. This effect is predicted by the work of Shannon which shows that all bandwidth expansion systems which attempt to achieve an improved signal-to-noise characteristic by fully utilizing a greater bandwidth will exhibit a threshold characteristic. Above threshold the output signal-to-noise ratio can be increased at the expense of increased bandwidth; however, this results in an increase in the total noise power at the input to the discriminator and the threshold occurs at a proportionately higher carrier level.

A number of demodulation techniques for improving the threshold performance of FM systems have been proposed and some of these techniques have been shown experimentally to exhibit substantially better performance

than the conventional discriminator. However, it should not be surprising that the discriminator (designed simply to be the inverse of the modulation process) is not the optimum FM demodulator, because consideration has not been given to the a priori knowledge of the statistics of the modulation or of the additive noise.

2. Maximum A Posteriori Reception of FM Signals

The basic demodulation problem is to obtain the best estimate of the modulation signal from an observation of the received signal plus noise utilizing all available statistical data. In a previous work[1] application of statistical estimation theory yielded an expression for the solution to the FM demodulation problem which provides the a posteriori (after utilization of all available data) most probable estimate of the modulating signal. The theory of maximum a posteriori (MAP) demodulation was first developed by Youla[3] and applied to amplitude and phase modulated systems. In our previous effort the application of this theory to frequency modulated systems was developed and the salient results of this work are summarized below.

The aim of this approach was to define a demodulator, which after having observed the received signal over some interval, T, produces as its output a function $\hat{a}(t;T)$, which is the MAP estimate of the transmitted modulation, $a(t)$.

For FM, the received signal is

$$e_R(t) = E_0 \sin\left[\omega_0 t + \beta\int_{t-T}^{t} a(u)\,du + \theta_0\right] + n(t),\ t-T \le u \le t \qquad (4)$$

where the signal is observed over the preceding T seconds. The MAP estimator chooses $\hat{a}(t,T)$ such that $p\left[\hat{a}(t;T)\,\middle|\,e_R(\cdot)\right]$ is maximized.

It is assumed that both the intelligence, a(t), and the noise, n(t) are zero mean Gaussian processes. It is further assumed that a(t) is stationary and that the additive noise is white and Gaussian, with a noise power density of N_0 watts per Hz (one-sided spectrum). Then, considering the infinite delay case, the MAP formulation leads to the following integral equation whose solution is the desired estimate. (See equation (26) in Chapter II of Reference [1].)

$$\hat{a}(\tau,\infty) = \frac{E_0\beta}{N_0}\int_{-\infty}^{\infty} h(\tau-x)\Big[E_0 \sin\beta\int_{-\infty}^{x}[a(u)-\hat{a}(u,\infty)]\,du$$

$$+ n_c(x)\cos\beta\int_{-\infty}^{x}\hat{a}(u,\infty)\,du - n_s(x)\sin\beta\int_{-\infty}^{x}\hat{a}(u,\infty)\,du\Big]\,dx$$

Let $n_1(x) = n_c \cos\hat{\theta}(x) - n_s \sin\hat{\theta}(x)$

where $\hat{\theta}(x,\infty) = \beta\int_{-\infty}^{x}\hat{a}(u,\infty)\,du$

Since n_c and n_s are independent white Gaussian noise processes of intensity $2N_0$, $n_1(x)$ is also a white Gaussian noise of intensity $2N_0$ (single-sided spectrum). The integral equation to be solved can now be written

$$\hat{a}(\tau,\infty) = \frac{E_0\beta}{N_0}\int_{-\infty}^{\infty} h(\tau-x)\Big[E_0 \sin\beta\int_{-\infty}^{x}\{a(u)-\hat{a}(u,\infty)\}\,du + n_1(x)\Big]dx\ . \qquad (6)$$

It is noted that the above equation is in the form of a convolution integral. From linear network theory it is known that the output of a linear filter is obtained as the convolution of the input signal and the impulse response of the filter. That is, if $g(x)$ is the input to a linear filter having an impulse response $h(x)$, the output, $y(\tau)$ is given by

$$y(\tau) = \int_{-\infty}^{\infty} h(\tau-x)\ g(x)\,dx\ . \qquad (7)$$

Thus, one may think of the terms within the square bracket of Equation (6) as being the input to a filter whose impulse response is $h(\tau)$ and whose output would be $\hat{a}(\tau,\tau)$.

Letting $\theta(x) = \beta\int_{-\infty}^{x} a(u)\,du$ and $\hat{\theta}(x,\infty) = \beta\int_{-\infty}^{x}\hat{a}(u,\infty)\,du$, Equation (6) may be represented by the following model for an infinite delay MAP FM demodulator.

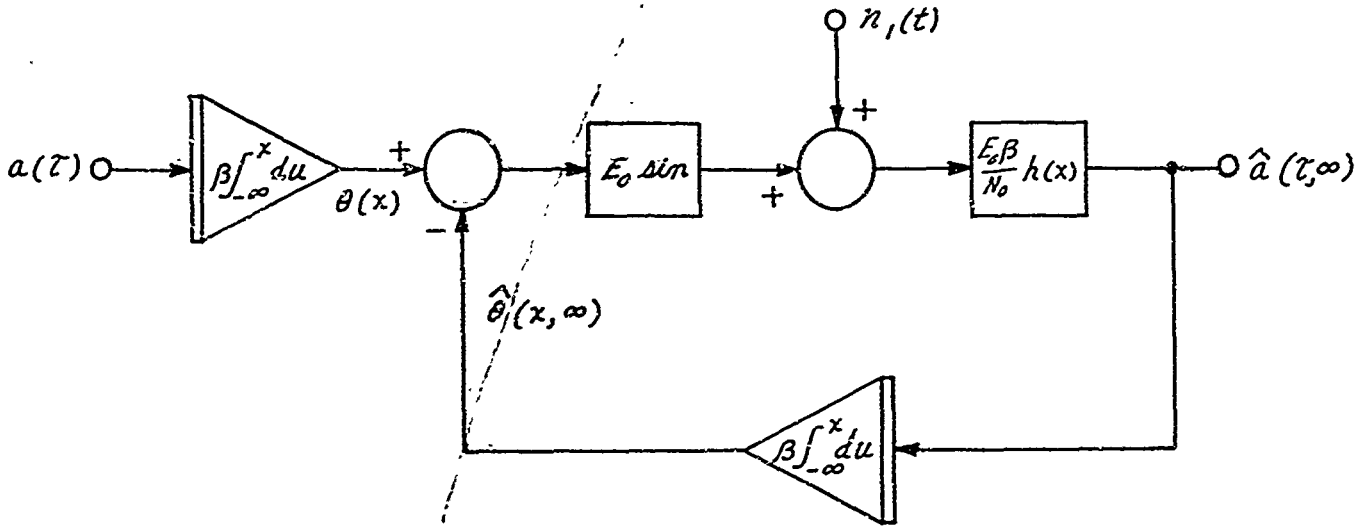


Figure V-2 MODEL FOR MAXIMUM LIKELIHOOD FM DEMODULATOR

The model in Figure V-2 is of the same form as the low frequency models which have frequently been employed to represent phase-locked loop receivers. A significant result presented in Chapter III of Reference [4], which relates the design of a PLL loop structure to the results of the MAP analysis, shows that if the model in Figure V-2 is linearized (i.e., if the sine operator is replaced by a unity gain), the loop filter as required for the MAP model is identical to the filter in the infinite delay phase-locked loop model. In the case of the phase-locked loop, the optimal filter is obtained by applying linear least mean-square filter theory to a linearized version of the loop. In this case the filter is known to be optimum in the MMSE sense only when the signal-to-noise is sufficiently high so that the linearized model is valid. However, from the MAP model, this same filter is also specified for MAP estimation when the sinusoidal operator is retained in the model. That is, it has been shown that an FM demodulator using a phase-locked loop structure containing this nonrealizable filter would perform the operation specified for the infinite-delay MAP demodulator. This equivalence is valid for all modulation spectra and at all carrier-to-noise ratios. It must be emphasized that the resulting structure is nonrealizable. This nonrealizable structure, however, suggests the design of realizable demodulators which hopefully would yield performance close to that of an optimum realizable demodulator.

The analysis presented in Chapter II of this report shows that the MAP estimate agrees with the series expansion for the MMSE estimate up to the first two terms. The empirical investigations reported in Chapter IV show close agreement of the MAP and MMSE estimates above and into the threshold region.

In a realizable phase-locked loop, the impulse response $\mathring{h}(x)$ must be zero for $X < 0$. Therefore, a realizable phase-locked loop would have a transfer function which differs from that required of an infinite delay MAP receiver. This difference in response could, in principle, be corrected in a linearized model by following the phase-locked loop by a compensating filter having the required transfer function. However, it is not clear that this would represent the optimum configuration for the nonlinear model at all carrier-to-noise ratios and thereby represent the system with the best threshold characteristic.

Similarly, referring to the MAP model, it has not been established that the realizable model in which the filter is designed as a zero-delay least mean-square filter will represent the best system at the lower carrier-to-noise ratios where the nonlinearity due to the sine operator will cause a distinct threshold characteristic. However, it does seem reasonable to accept this design approach as an approximation to a MAP demodulator and then compare the resulting performance with other systems. Because it is very difficult to determine the resulting performance in the threshold region by analytical means, the low frequency equivalent circuit, Figure V-3, was used in the initial phase of an analog computer simulation and is representative of a zero-delay PLL structure.

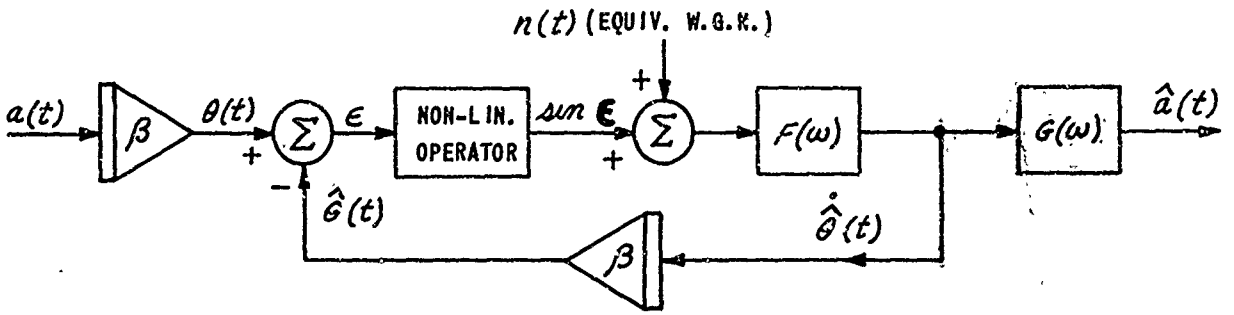


Figure V-3 LOW-FREQUENCY EQUIVALENT CIRCUIT OF PLL FM DEMODULATOR

3. The Phase-Locked Loop Simulation

The computer mechanization of the low frequency equivalent circuit in Figure V-3 was straightforward. Some of the important considerations regarding selection of the modulation function, definition of signal bandwidths, synthesis of the loop and post loop filters, and mechanization of the sinusoidal nonlinearity are discussed below.

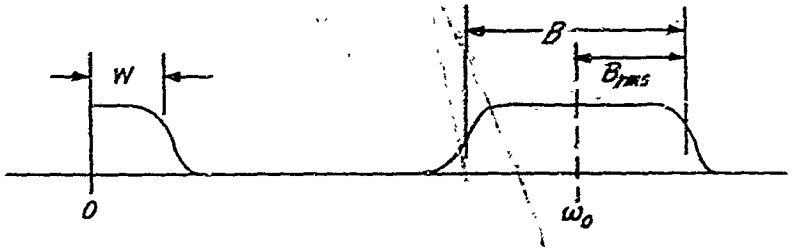
Both random and sine wave modulating signals, a(t), were used in this simulation. The random modulation function was generated by passing white Gaussian noise through a third order Butterworth filter. The resulting power spectrum of the random modulation signal is given by

$$S_a(\omega) = \frac{A_o^2}{1+\left(\frac{\omega}{\omega_c}\right)^6} \tag{8}$$

Our earlier theoretical work[1] on the MAP estimator for FM employed a random (Gaussian) modulation function which, for mathematical convenience corresponded to the use of a first order Butterworth filter. However, it was felt that a modulation spectrum which falls off at this rather slow rate is not representative of the modulation usually encountered in practice, therefore a sharper cutoff spectrum was chosen for this simulation. In order to avoid any possible error due to D.C. drifts in the associated equipment, the low frequency end of the modulation spectrum was removed by means of a high pass filter. The frequency of the sine wave modulation was chosen equal to the corner frequency, ω_c , of the third order Butterworth filter.

With each form of modulation it is necessary to establish a somewhat arbitrary, although quantitative, value for the RF bandwidth of the FM signal and then define a "bandwidth expansion factor", M, which is the ratio of the RF bandwidth to the modulation bandwidth. A useful result which relates the rms bandwidth of the modulated signal to the rms bandwidth of the modulating signal when the modulation is a stationary Gaussian random process, has been derived by Abramson.[5] Abramson finds that for a frequency modulated wave the rms bandwidth (in Hz) of the modulated wave is $1/2\ \pi$ times the rms frequency deviation in radians/second. The bandwidth as defined by Abramson is the rms bandwidth referenced about the

carrier frequency. It would seem that a reasonable measure of the required RF bandwidth would be twice the rms value as shown in the Figure below.



Then the bandwidth expansion factor (M) is defined as the ratio of twice the rms bandwidth of the modulated wave to the rms bandwidth of the modulating wave. This brings the definition of bandwidth into close agreement with Carson's rule. Consider a carrier the frequency of which is sinusoidally deviated with a maximum deviation of Δ f Hz at a modulating frequency f_m Carson's rule, which defines the required bandwidth and hence corresponds to the width of a zonal filter passing the significant energy components of the modulated signal, yields,

$$B_c = 2 f_m \left(\frac{\Delta f_{max}}{f_m} + 1\right) = 2 f_m (\mu + 1) \tag{9}$$

At large values of μ, B_c approaches twice the value of the peak deviation. Defining the RF bandwidth as twice the rms bandwidth yields

$$B = 2 B_{rms} = 2 \Delta f_{rms} \tag{10}$$
$$= \sqrt{2}\, \Delta f_{max} = \frac{B_c}{\sqrt{2}}$$

which agrees with Carson's rule in the limit of large modulation index recognizing that B_{rms} is equal to the rms deviation rather than the peak deviation. When Carson's rule is applied to other than sinusoidal modulation, f_m denotes the highest frequency component in the modulation. The bandwidth expansion factor is then $2\left(\frac{\Delta f_{max}}{f_m} + 1\right)$

When sinusoidal modulation was used for the testing of FM systems, the frequency was chosen equal to the 3 db frequency of the Butterworth filter and the bandwidth expansion factor was considered to be given by $2(\mu + 1)$.

Mechanization of the required sinusoidal nonlinearity presented the greatest challenge, and was successfully achieved through the use of a modified rate resolver circuit in which a function generator provided the basic nonlinear characteristic. In its normal application the input to the function generator must be restricted to values between $\pm$ 180°; however, this range was extended by the use of automatic switching, controlled by digital logic in the hybrid computer. This permitted operation of the rate resolver circuit over the large values of the phase error encountered in the threshold region. In mechanizing the sinusoidal nonlinearity with the rate resolver circuit configuration, the derivative of the error signal is required as the input, and therefore, a(t) and $\dot{\hat{\theta}}(t)$ may be used as inputs to the summing amplifier. Thus, the two integrators in Figure V-3 could be eliminated.

The filters were designed such that the loop was optimized as a minimum mean-square error system using the linearized model, i.e., setting $\sin \epsilon(t) \doteq \epsilon(t)$. First, the loop filter, F(s), was designed to minimize the mean-square phase error $\langle \epsilon^2(t) \rangle$ of the linearized loop. Then, with F(s) specified, G(s) was chosen to minimize the mean-square error between the modulation function a(t) and the demodulator output, $\hat{a}(t)$. The details of the synthesis of these filters and the evaluation of the theoretical mean-square error for the linearized system are given in Appendix V-A.

As noted previously, the low frequency equivalent circuit of a realizable zero-delay PLL can be made equivalent to the linearized model of Figure V-2 by incorporating a post loop filter with delay. The threshold phenomenon is associated with the loop characteristics and therefore, the design of the post loop filter will not influence the threshold. For convenience in measuring the MSE, the post loop filter G(s) was designed to produce the MMSE zero-delay estimate of a(t). However, the above-threshold performance will be poorer than that obtained from a closer approximation to the infinite-delay model. It is again emphasized that the primary intent of this simulation effort was to explore the threshold characteristics which are difficult to obtain analytically, whereas the theoretical performance above threshold is obtainable by straightforward analytical techniques, viz Chapers II and IV.

4. Experimental Results

The initial set of threshold curves obtained in this simulation were made with random modulation and bandwidth expansion factors of 20 and 80. Experimental data was taken with and without the nonlinear sine operator in the loop. This permitted a useful check with the theoretical evaluation of the above threshold mean-square error. The theoretical performance of the linearized system for a bandwidth expansion factor of 80 is shown by curve 1 in Figure V-4. The corresponding experimental data from the simulation of the linearized model are shown as curve 2. The experimental data is seen to be in excellent agreement with the computed theoretical performance. The loop filter and post loop filter were optimized at each value of input carrier-to-noise ratio to achieve the minimum mean-square error performance of the linearized model. These same filters were then employed when the nonlinear operator was placed in the circuit to represent the actual PLL system. The threshold characteristic obtained under these conditions is shown by curve 3. It seems reasonable that the performance below threshold could be improved somewhat by increasing the loop gain over that specified by the linear optimization. This is suggested by the quasi-linear analysis[4] of such loops where the nonlinear operator is replaced by an equivalent linear gain. The results of this analysis show that at the lower carrier-to-noise ratios where the loop phase error is sufficient to reduce the gain of the equivalent linear operator (which is made equal to the average gain of the nonlinear element), the gain of the filter should be increased to compensate for this decreased average gain. Gain adjustments were made on an empirical basis and the threshold characteristic shown by curve 4 was obtained. This curve shows that an improvement of approximately 1 db in (S/N), at which threshold occurs is obtainable by this technique.

The loop threshold characteristic was also investigated in the absence of a modulation signal but with the loop designed on the assumption that the modulation is present and optimized in the same manner as for curve 4. Above threshold, where the linearized model of the system is valid, the optimum linear filtering results in a minimum mean-square error which is divided between a distortion term due to the filtering action of the loop and a noise term due to additive noise. Therefore, with the modulation removed,

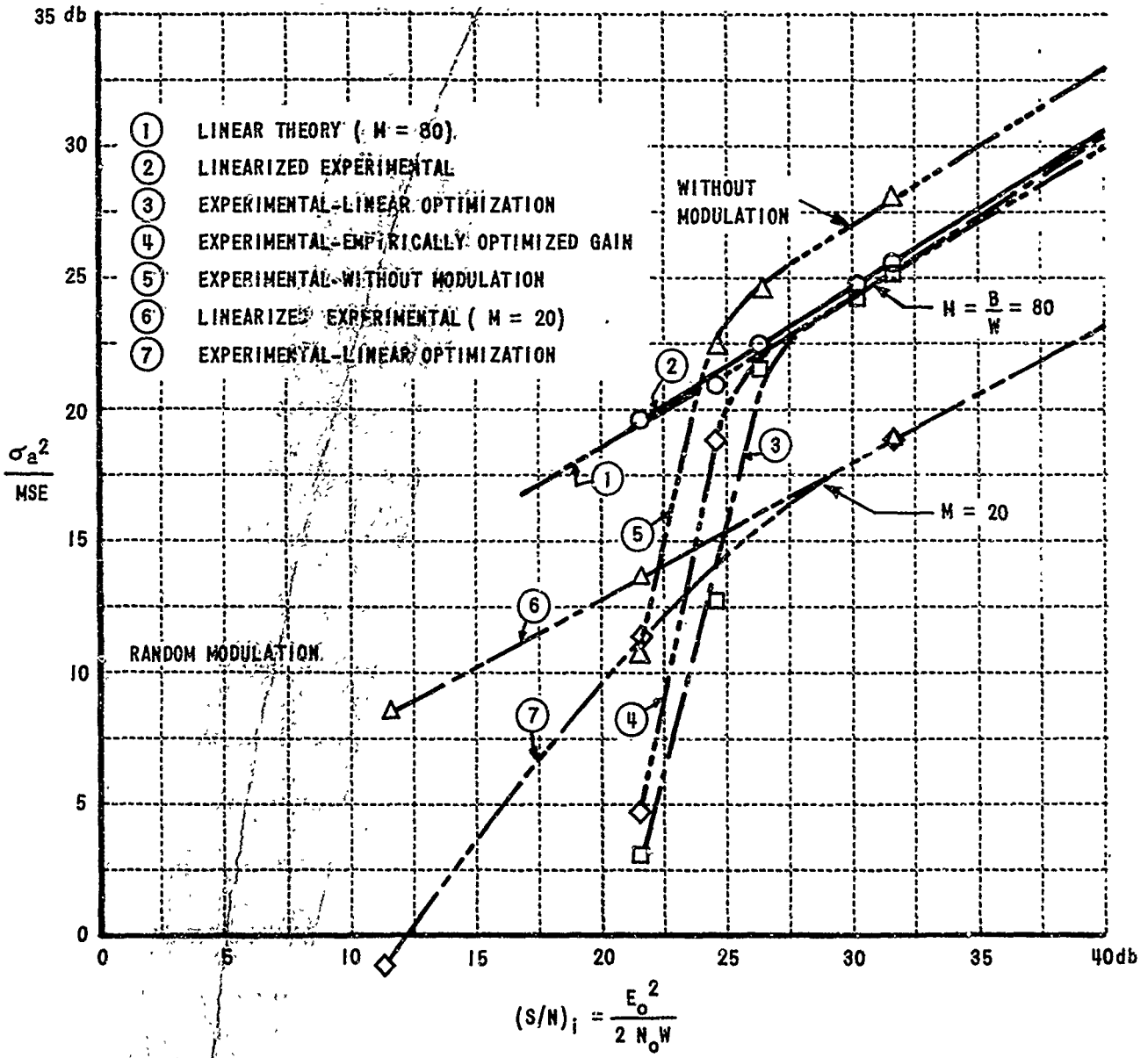


Figure V-4 SIMULATION OF PLL FM DEMODULATOR

a decrease in the mean-square error would be expected. (In this case the MSE = $\sigma_{\hat{a}}^2$ since $\sigma_a^2 = 0$; however, for purposes of comparison, σ_a^2 /MSE was computed as if σ_a^2 had retained its full value.) The experimental data is plotted in curve 5 and the expected reduction in the mean-square error above threshold is observed. With this particular loop design the experimental data shows virtually no change in the input carrier-to-noise ratio at which threshold occurs with or without modulation.

Curves 6 and 7 show the experimental data obtained with random modulation and with a bandwidth expansion factor of 20. Here, we observe that the threshold curve for the linearly optimized system is much more gradual than when the larger bandwidth expansion factor was employed. In fact, this curve could be raised even further by empirical adjustment of the loop gain in the region below threshold.

It might be noted that the curves above threshold do not have the unity slope which usually depict the performance of practical demodulators. The usual curves apply to fixed filter systems and consider only the error in reproduction due to the noise which decreases inversely with input signal-to-noise ratio. The error caused by distortion of the modulation function (resulting from the filtering action of the demodulation system) is neglected. Ideally, the unity slope would be obtained if the system had been designed for and used with a modulation function which was strictly bandlimited. However, the curves, shown in Figure V-4, were obtained for a system with the modulation spectrum as given by Equation (3) and the loop filters were adjusted as a function of the carrier-to-noise ratio in order to maintain the minimum mean-square error condition taking into account both distortion and noise components. That is, with increasing carrier-to-noise ratios, the system bandwidth is increased to accept more signal energy; otherwise the distortion term would soon dominate over the noise term and, in fact, the ordinate would no longer rise linearly but would quickly approach a limit set by the distortion term.

The performance of the PLL, designed for this random modulation spectrum, was also evaluated when sinusoidal modulation was used. In this case, the required RF bandwidth is more easily defined and comparisons can

be made with generally accepted FM (discriminator) curves and the results of other reported PLL investigations for which sinusoidal modulation was assumed or employed.

In our simulation, the frequency of the sinusoidal modulation was set at the corner frequency, ω_c, of the random modulation spectrum given by equation (8). The bandwidth expansion factor, M, was defined by Carson's rule to be

$$M = 2(\mu+1)$$

It was observed that about the same performance above threshold, i.e., about the same value of σ_a^2/MSE for a given input carrier-to-noise ratio, was obtained when the mean-square value of the sinusoidal modulating signal was made equal to the mean-square value of the random modulation function. This also results in approximately the same bandwidth expansion factor, and therefore, the system was "optimized" as for the case of the random modulation function with the equivalent bandwidth expansion factor. The filters were identical to those designed for use with the random modulation function.

The threshold curve in Figure V-5 shows the performance of the PLL system when a sinusoidal modulation with a modulation index of 40 was employed. The loop transfer function was optimized (with the empirically) established optimum loop gain below threshold) as described previously.

The threshold is seen to occur at an $(S/N)_i$ of approximately 24 db. In this case, the bandwidth expansion factor is approximately 82 and the curve would have to be shifted 19 db to the left if the abscissa scale indicated the carrier-to-noise power ratio in the RF bandwidth. On that basis, the threshold occurs at 24-19 = 5 db which is significantly better than the 12 db carrier-to-noise ratio at which threshold occurs in the conventional discriminator type demodulator.

Figures V-6, V-7, and V-8 show signal recordings obtained during the simulation which resulted in the threshold curve of Figure V-5. These three figures show typical waveforms taken just above threshold, at threshold, and below the threshold point, respectively. In each Figure, curve A is the modulation function, a(t) (sinusoidal in this case) and is the input to the low frequency equivalent circuit used in this simulation. Curve B is the estimate

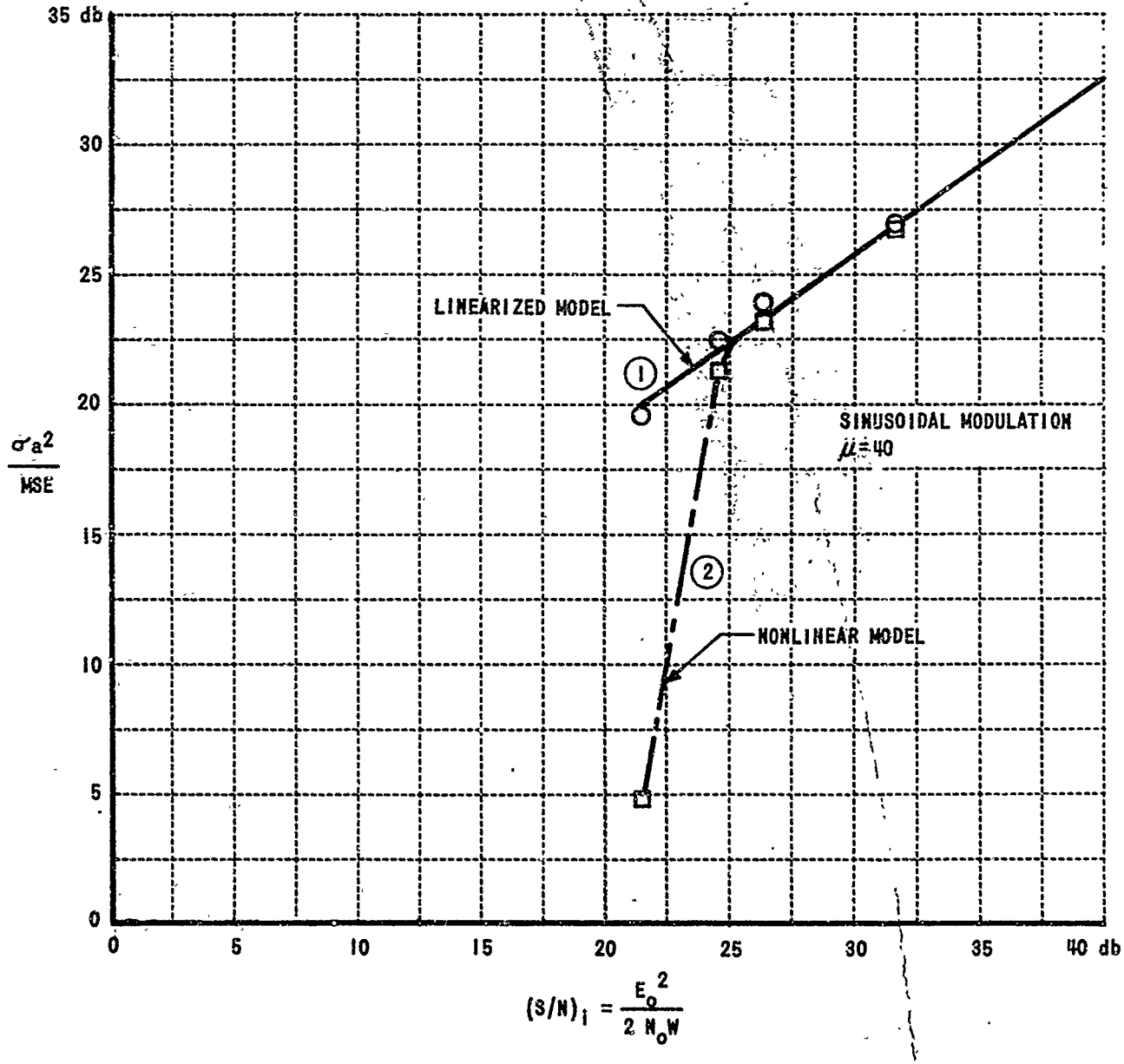


Figure V-5 SIMULATION OF PLL FM DEMODULATOR-SINUSOIDAL MODULATION

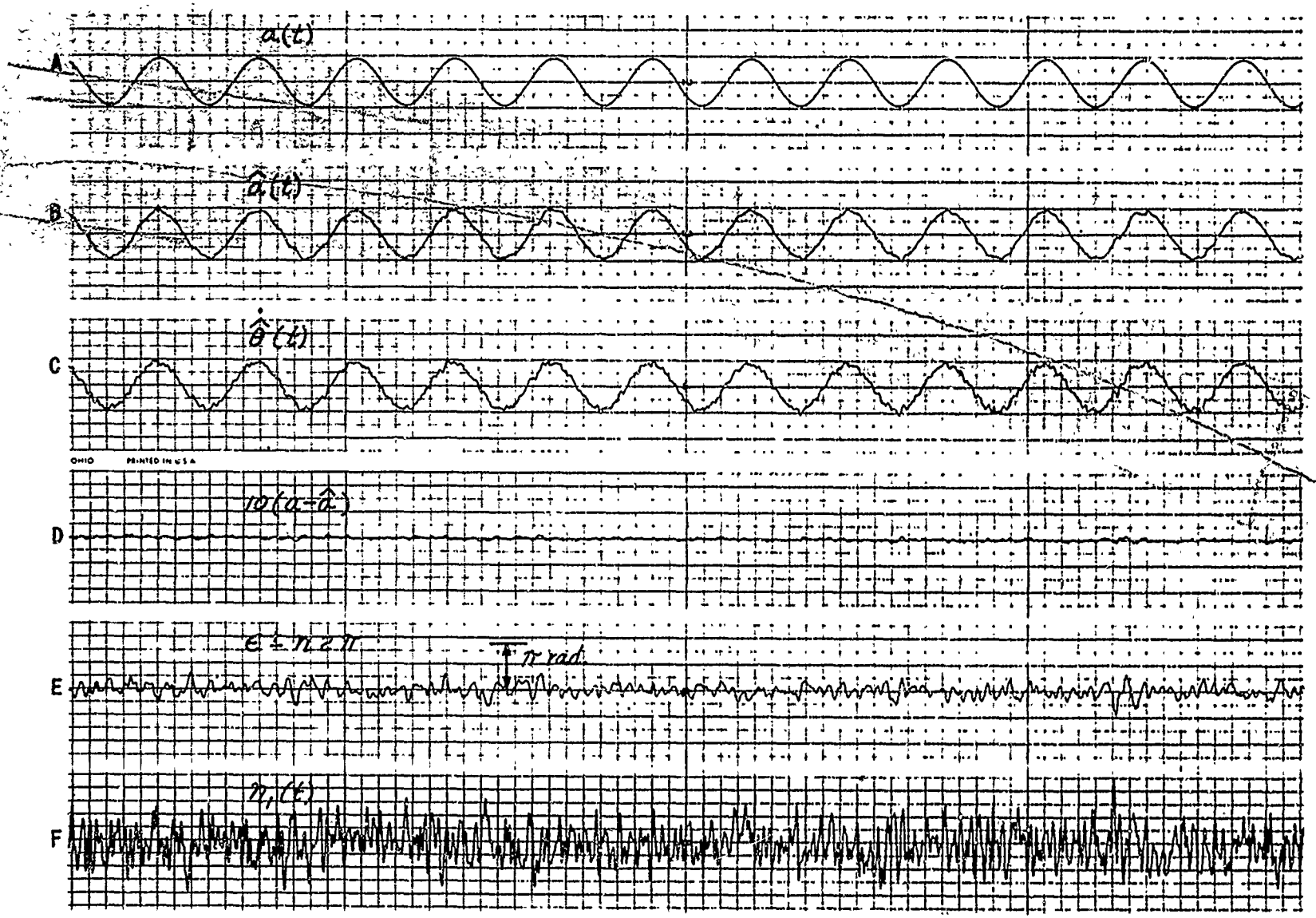


Figure V-6 RECORDINGS FROM SIMULATION OF PLL FM DEMODULATOR ABOVE THRESHOLD
$(S/N)_i$ = 26.5 db , μ = 40

Figure V-7 RECORDINGS FROM SIMULATION OF PLL FM DEMODULATOR AT THRESHOLD
$(S/N)_i$ = 24.5 db , μ = 40

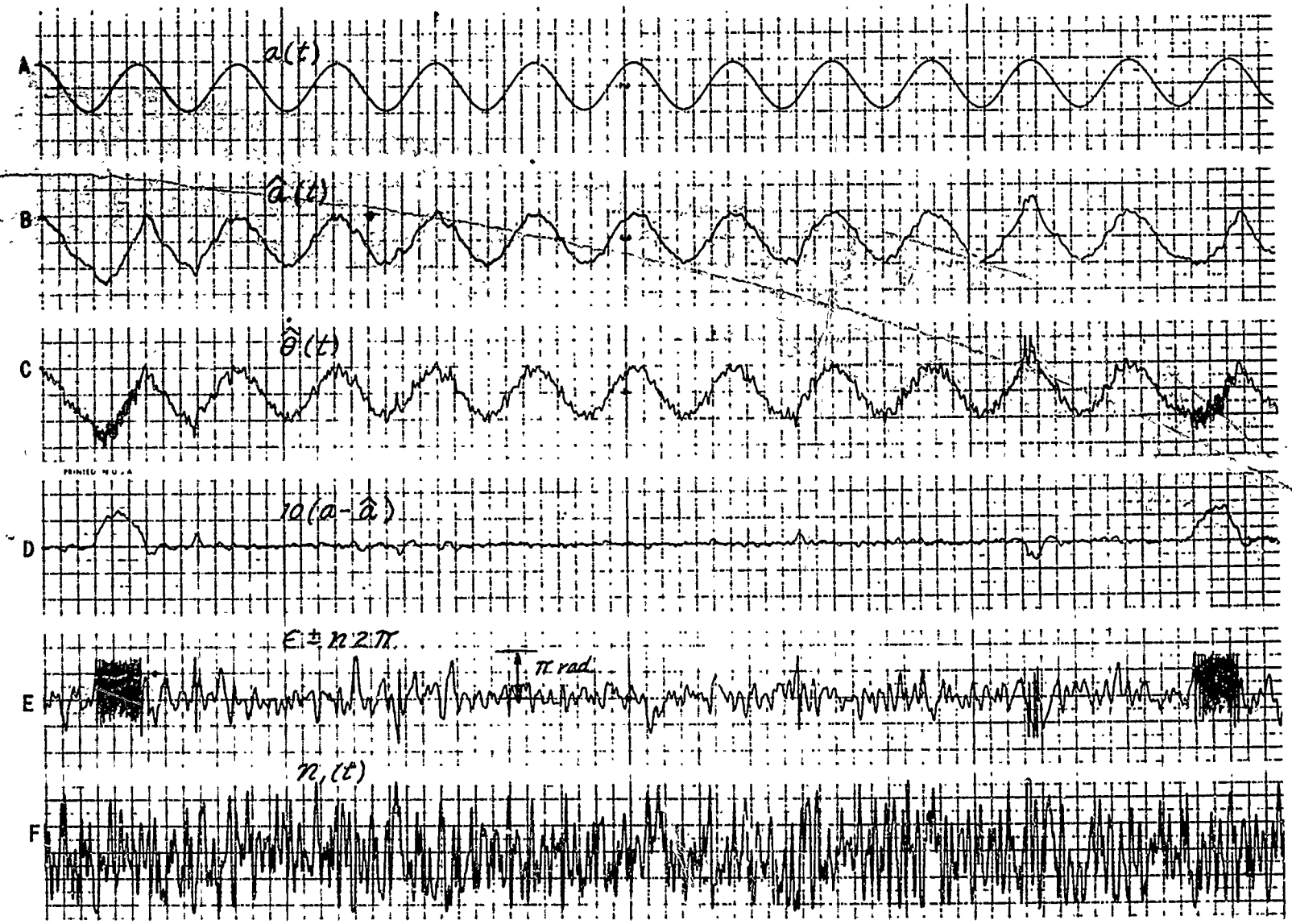


Figure V-8 RECORDINGS FROM SIMULATION OF PLL FM DEMODULATOR BELOW THRESHOLD
$(S/N)_i$ = 21.5 db , μ = 40

of the modulation function produced at the output of the post loop filter of Figure V-3. The signal at the output of the wideband PLL is shown in curve C. This is the signal actually fed back to the VCO in the physical system and then filtered by the post loop filter to form the estimate of $\hat{a}(t)$. Curve E is the error between the modulation function and the receiver estimate, i.e., the difference between curves A and B. Curve E shows the phase error signal at the input to the nonlinear operator in the equivalent circuit. This input is maintained through the digital logic within $\pm \pi$ radians and discontinuities in curve E indicate the number and direction of the skipped cycles. Curve F is a recording of the equivalent Gaussian noise introduced into the loop and this same sample of noise, recorded on tape, was used in each run.

In Figure V-6 the loop was operating above threshold at a $(S/N)_i$ of 26.5 db (see Figure V-5). The estimate is seen to closely reconstruct the original modulation function and no cycle skipping occurred during this run.

Figure V-7 shows waveforms taken at threshold with the $(S/N)_i$ set at 24.5 db. The estimate of the modulation (curve B) is seen to be somewhat more noisy than in the previous figure although it still appears to be a reasonably good estimate of a(t) most of the time. However, cycle skipping occurred during this run and two instances where loss of lock occurred are shown near the beginning and the end of the short sample of the recording reproduced in Figure V-7. These instances of cycle skipping are easily identified on Curve E in which the phase error is abruptly shifted $\pm 2\pi$ radians at each instant that the actual phase error exceeds $\pm 180°$, thus keeping the error signal recorded on Curve E within the bounds of $\pm \pi$ radians.

We note that the system skipped one cycle, relocked again and tracked along until two 2π skips occurred. It is also seen that these skips are coincident with the occurrence of large instantaneous values of the additive Gaussian noise shown in curve F. The resulting "click" phenomenon is also observed in the demodulator output in curve C, with the corresponding increase in the error as shown in curve D.

Waveforms resulting from operation below threshold at a $(S/N)_i$ of 21.5 db are shown in Figure V-8. Both single and multiple skips are seen to occur and the mean-square error performance has been degraded considerably.

5. Threshold Investigation of a Wideband PLL Design

The low frequency equivalent circuit of a PLL as used thus far in the simulation was modified to agree with the loop design employed in a study of the threshold of phase-locked loops at the Polytechnic Institute of Brooklyn[6]. In their investigation, a sinusoidal modulation function was assumed and data was obtained with and without modulation. The resulting threshold characteristic was considerably better than the generally accepted performance of a PLL demodulator and it was concluded that a wideband, critically damped loop (as opposed to a narrow band loop) was required to achieve this performance.

Only sinusoidal modulation was employed in this portion of our simulation. An ideal zonal (low-pass) filter with a cutoff frequency equal to the modulation frequency was assumed in the theoretical PIB model. For practical reasons, the post loop filter in our simulation was a third order Butterworth filter with the corner frequency equal to the modulation frequency of 10 radians per second. The design of the loop filter, for a modulation index equal to 10, was carried out in accordance with the PIB report. The damping coefficient of the closed loop transfer function was adusted for maximum flatness, which, incidentally, results in a second order Butterworth filter. This procedure does indeed result in a loop bandwidth which is wide compared to the modulation frequency. In fact, this design resulted in a loop bandwidth approximately 53 times the modulation frequency with the loop designed for a modulation index equal to 10. This design approach appears to be in general agreement with the loop design which results from the minimum mean-square error approach employed in the previous phase of this simulation. That is, both approaches lead to a loop bandwidth which is much wider than the bandwidth of the modulation signal.

The simulation of the PIB PLL was carried out in much the same way as the other PLL simulation described earlier in the chapter. However, in this particular model a sharper cutoff post-loop filter was employed in order to be in close agreement with the PIB model. As a result, a significant phase shift was introduced by the post-loop filter at the modulation frequency. A satisfactory mean-square error measurement was easily made near and below

threshold; however, at very high output signal-to-noise ratios above threshold the total mean-square error was dominated by distortion of the modulation signal.

In the absence of modulation, the output is due entirely to noise and represents the total mean-square error (with respect to the modulation, which by definition is set equal to zero). The output signal-to-noise ratio is plotted in terms of $\frac{\sigma_a^2}{MSE}$ computed as if a modulation signal of power σ_a^2 was present. The origin of the ordinate has been adjusted so as to bring the experimental results, which were recorded only on a relative basis, into agreement with the known asymptotic results at high carrier-to-noise ratios. The value of the intrinsic input carrier-to-noise ratio, $(S/N)_i$, at which threshold occurs, is quantitatively obtained from the abscissa.

The threshold curves of Figure V-9 show the threshold performance without modulation for five different loop bandwidths from 50 to 530 radians per second. The post-loop filter had its -3 db point at the modulation frequency of 10 radians per second. When the modulation signal was added in the narrow bandwidth loop, the loop was observed to skip cycles even in the absence of noise. The noise-free skipping could not be eliminated until loop bandwidths of 200 radians per second or higher were employed.

Figure V-10 shows the threshold curves with and without modulation when the loop bandwidth is adjusted to 200 radians per second. In this case, the threshold occurs at approximately 5 db greater $(S/N)_i$ when the modulating signal is present.

In Figure V-11 with the wideband loop, we see that the influence of the modulation function is much less (approximately 1 db) than with the narrower loop bandwidth systems. However, it was observed that the threshold with modulation was not a very sensitive function of the loop bandwidth, and in fact, the threshold is seen to occur at approximately the same input carrier-to-noise ratio with either of the two largest loop bandwidths.

As a result of this simulation we find the threshold occurs at somewhat higher carrier-to-noise ratios than indicated in the PIB report and appears to be consistent with most other PLL results which indicate that a significant

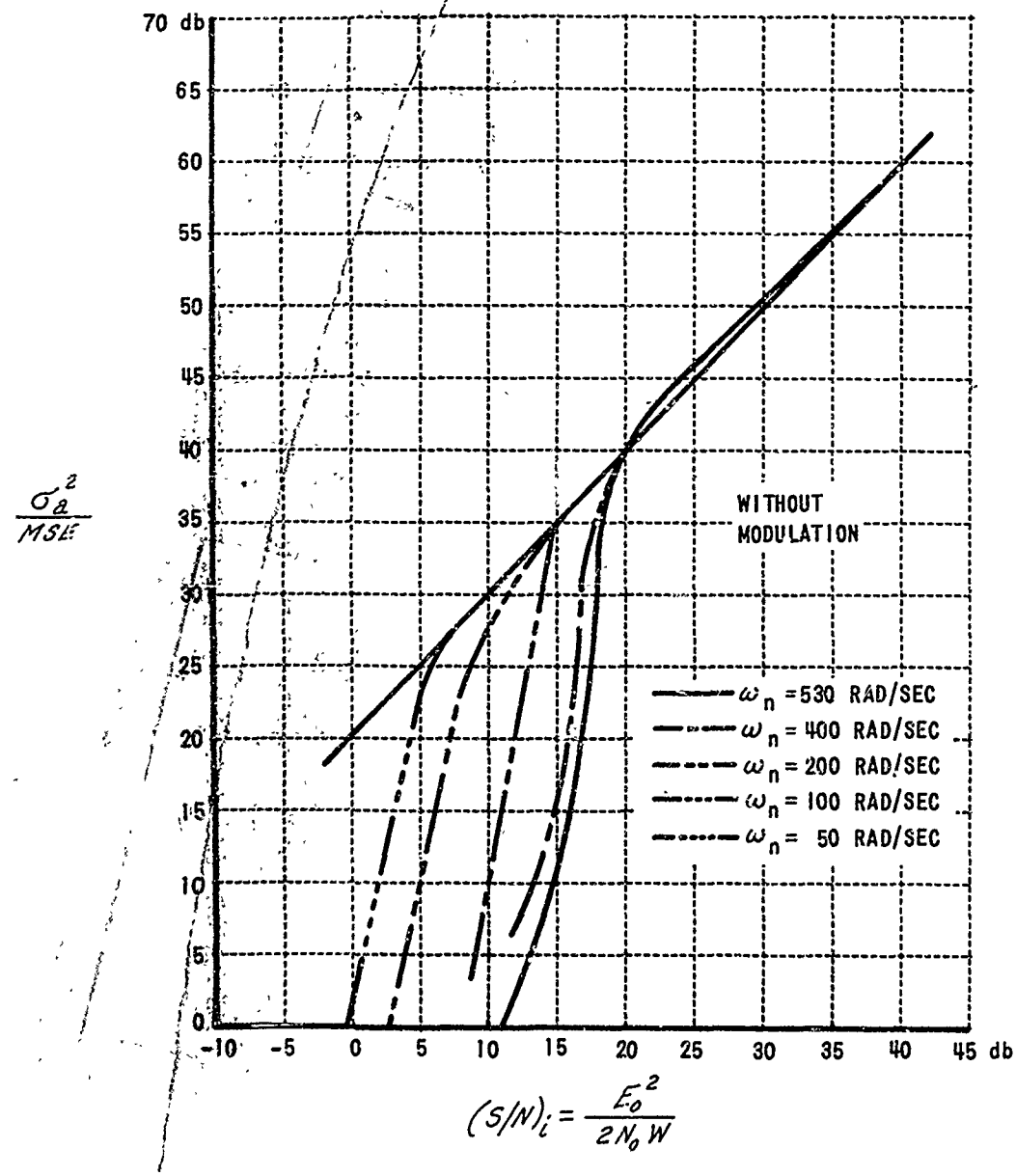


Figure V-9 SIMULATION OF PLL PERFORMANCE AS A FUNCTION OF LOOP BANDWIDTH

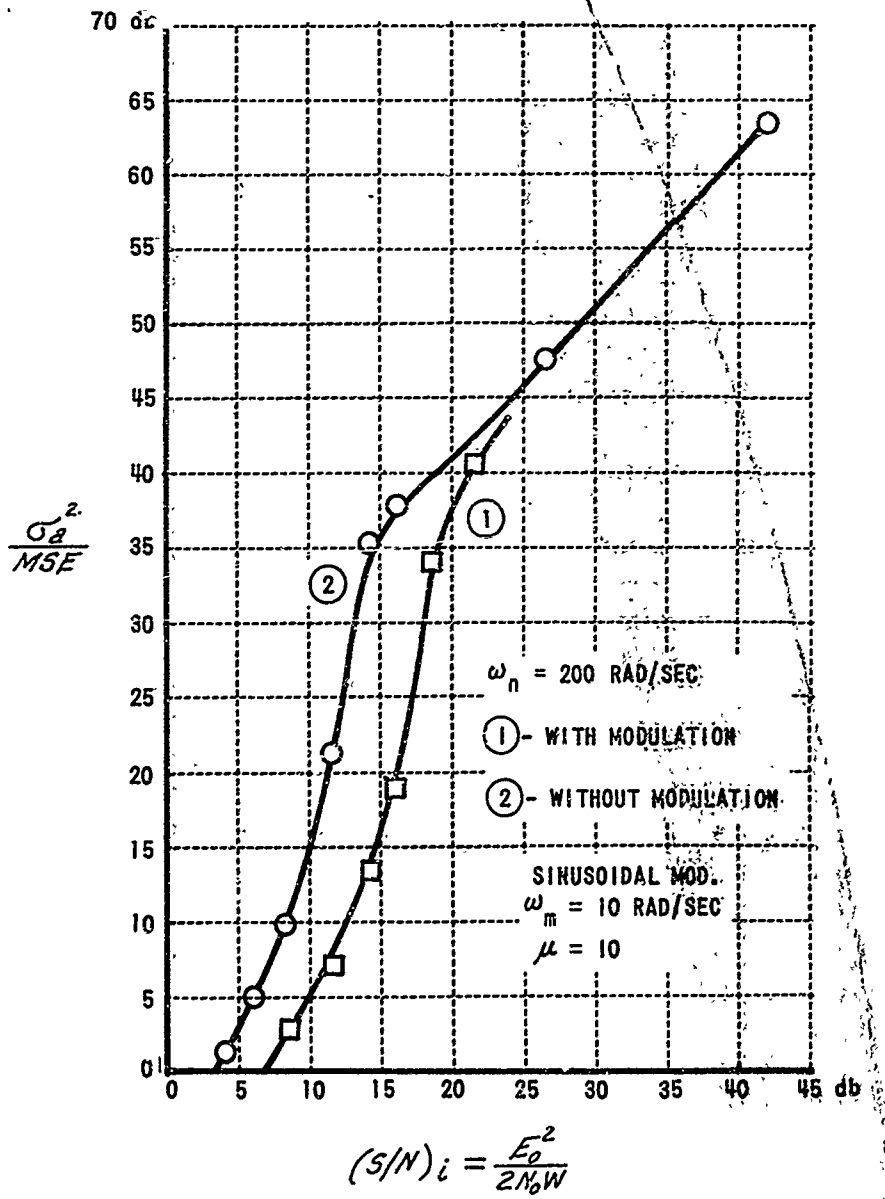


Figure V-10 SIMULATION OF PLL FM DEMODULATOR-EFFECT OF MODULATION

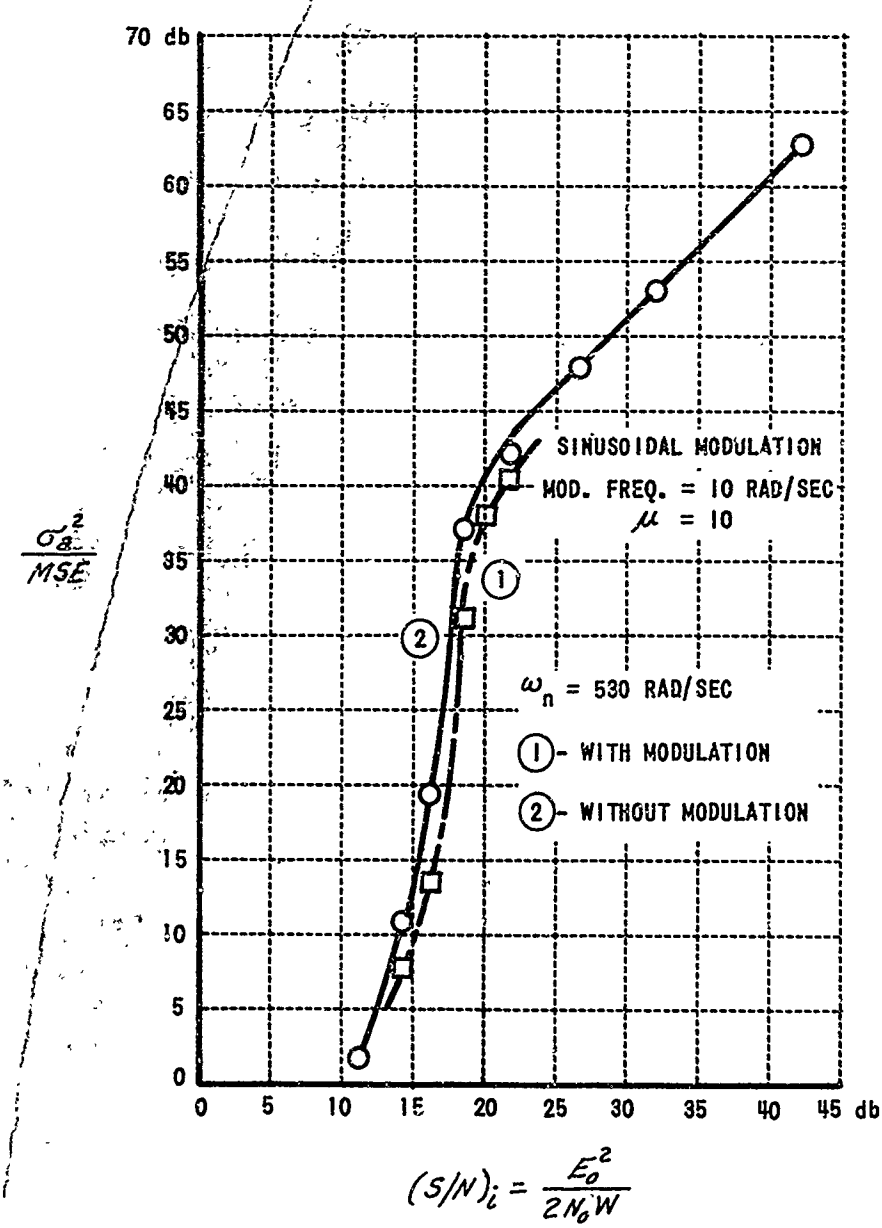


Figure V-11 SIMULATION OF WIDE-BAND PLL FM DEMODULATOR

threshold improvement, with respect to a discriminator, can be achieved with conventional PLL systems, for the values of μ normally encountered.

6. Comparison with the Discriminator

It is interesting to compare the results of this simulation with the threshold characteristics of a conventional FM receiver employing a discriminator. Figure V-12 is a plot of the input carrier-to-noise power ratio at which threshold occurs as a function of the bandwidth expansion factor in an FM system. The data for the discriminator was taken from the well known curves derived by F. J. Skinner.[7] Two points from our simulation using a sinusoidal modulation signal are also shown. The lower curve is taken from a paper by Akima[2] in which he theorized how far the threshold extension can be carried in a realizable system. It is seen that the performance obtainable with the PLL demodulator is significantly better than the discriminator, but falls well short of the Akima limit. It should further be noted that other theoretical studies such as described in Chapters II and IV of this report suggest that the threshold may possibly be improved well beyond the Akima Limit, particularly if nonreal time processing is employed.

7. Conclusions

1) It is concluded that the use of a hybrid computer simulation is a useful and accurate technique for evaluating the threshold performance of PLL demodulators. This technique can be extended to investigate the performance of even more complex structures, such as a frequency demodulator with feedback when a PLL is employed as the frequency detector in the FMFB loop.

2) As noted earlier in this chapter, a previous effort[4] established the relationship between the MAP estimation of an FM signal and a closed loop configuration which required a nonrealizable loop filter. However, the potential threshold extension of an infinite-delay MAP estimator cannot be realized with a zero-delay PLL. Furthermore, there is no proof that the conventional PLL structure represents even a very close approximation to the best possible realizable system. It is suggested that this simulation approach be used to investigate modified PLL designs which may lead to

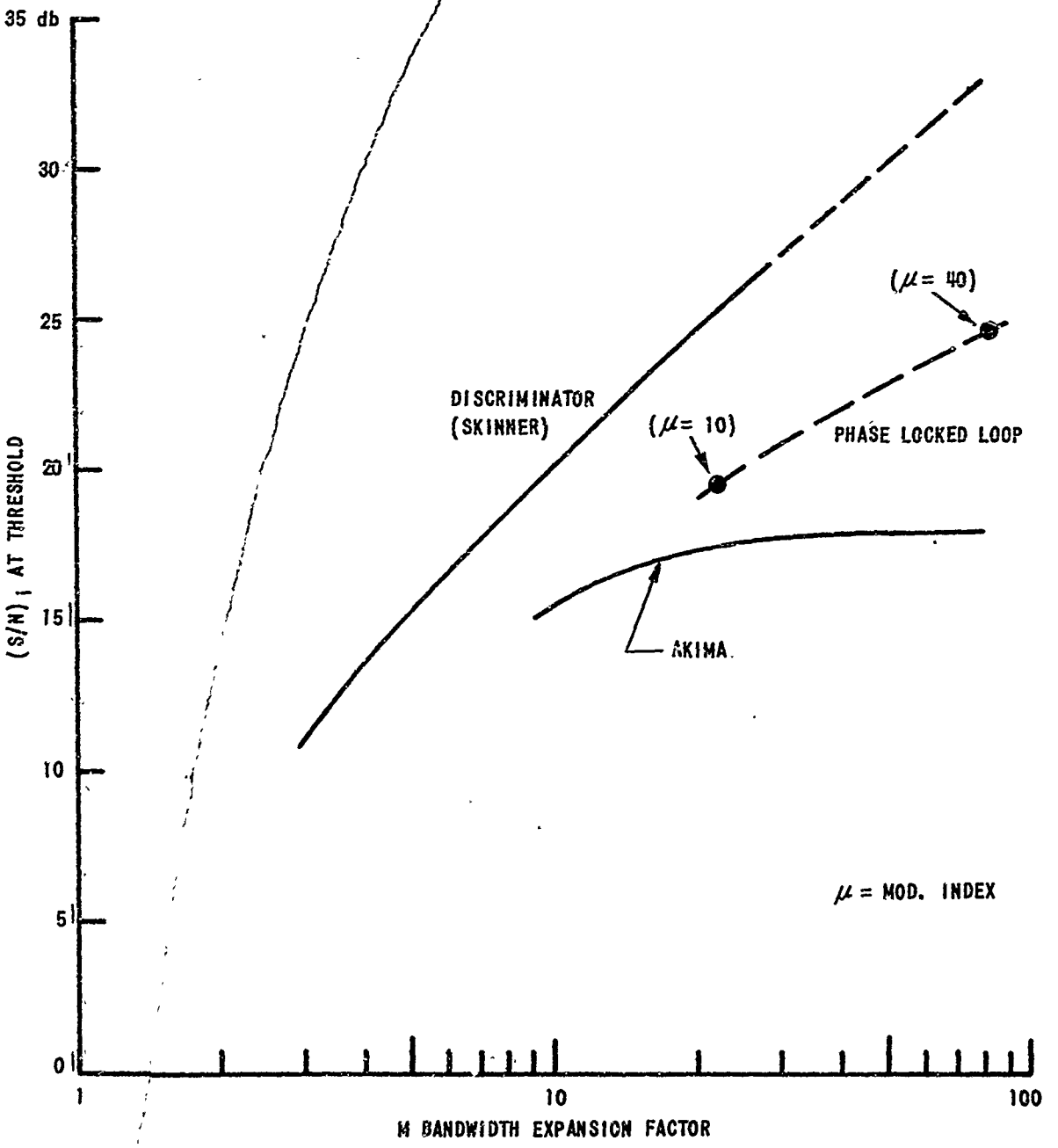


Figure V-12 COMPARISON OF FM DEMODULATOR THRESHOLD PERFORMANCE

improved realizable, zero-delay FM demodulators. Among these would be various PLL structures in which a modified phase detector characteristic is employed. This would include such techniques as Linlock[8], Tanlock[9], and the Extended Range Phase Detector (ERPD).[10] One should not restrict these efforts simply to modifications of the phase detector in a PLL system but consider improvements to the FMFB demodulator or other techniques that take advantage of the particular manner in which the additive noise and an FM signal interact.

3) A significantly better FM demodulator may be possible if the application does not require realtime processing. The operation specified by the infinite delay MAP analysis could be closely approximated if the received signal plus noise were available in stored form, as for example, by recording the I. F. signal. Demodulation might require computer processing and would entail some delay, at least of the order of the correlation time of the modulation, before the estimate of the modulation could be formed. In many applications the delay could be so small that these systems would be entirely practical in situations where "realtime" demodulators are ordinarily thought to be a necessity.

The intensive efforts made by many investigators over the past several years notwithstanding, there is no evidence that the limiting performance of angle modulated systems has been attained, and we anticipate that significant improvements in performance will be achieved. Two of the most promising approaches seem to be the use of a modified phase detector and the development of practical approximations to the "infinite-delay" MAP or MMSE estimators in situations where a small processing delay is tolerable.

APPENDIX V-A

SYNTHESIS OF THE LOOP AND POST LOOP FILTERS

The procedure for synthesizing the loop and post-loop filters of the low frequency equivalent circuit of Figure V-3 is described in this Appendix. In accordance with the design suggested from the MAP estimation theory, the transfer functions of the filters are obtained by applying least square error filtering theory to the linearized model shown in Figure V-1A. The loop filter F(s) is synthesized so as to minimize the mean-square phase error in the linearized loop. Then, with F(s) defined, the post-loop filter, G(s), is chosen to minimize the mean-square error between the modulation function a(t) and the demodulator output, $\hat{a}(t)$.

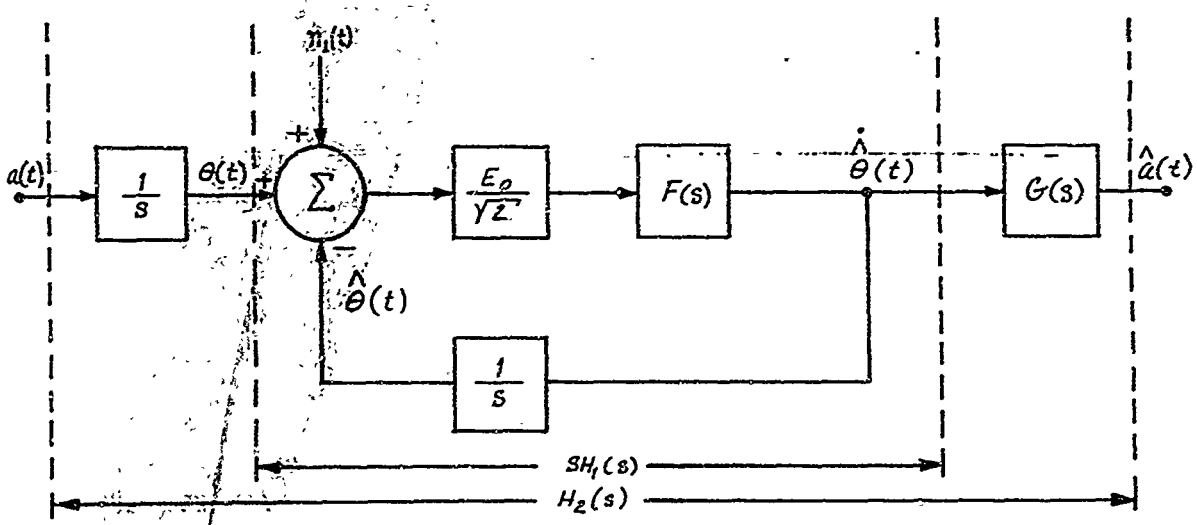


Figure V-1A

Let $\frac{\hat{\theta}(s)}{\theta(s)} = H_1(s)$ be the closed loop transfer function which minimizes the phase error, $\hat{\theta}(t) - \theta(t)$.

$$H_1(s) = \frac{\frac{E_0}{\sqrt{2}} \frac{F(s)}{s}}{1 + \frac{E_0}{\sqrt{2}} \frac{F(s)}{s}} \tag{A-1}$$

and

$$F(s) = \frac{\sqrt{2}}{E_0} \cdot \frac{s H_1(s)}{1 - H_1(s)} \tag{A-2}$$

Let $\frac{a(s)}{\hat{a}(s)} = H_2(s)$ be the transfer function which minimizes the difference between the modulation function a(t) and the demodulator output, $\hat{a}(t)$

Thus
$$H_2(s) = \frac{\frac{E_0}{\sqrt{2}} \frac{F(s)}{s}}{1 + \frac{E_0}{\sqrt{2}} \frac{F(s)}{s}} \quad G(s) = H_1(s) G(s) \tag{A-3}$$

$$G(s) = \frac{H_2(s)}{H_1(s)} \tag{A-4}$$

Initially in the experiment we used a random modulation function obtained by passing white Gaussian noise through a third order Butterworth filter resulting in a modulation spectrum

$$S_a(s) = \frac{A_0^2 \omega_c^6}{-s^6 + \omega_c^6} \tag{A-5}$$

The equivalent additive noise was obtained from a white Gaussian noise source, independent of the modulating signal, having a power spectral density

$$S_{n_1}(s) = \frac{2 N_0}{E_0^2} \tag{A-6}$$

The power spectral density of $\theta(t)$ is given by

$$S_\theta(s) = \frac{A_0^2 \omega_c^6}{-s^2(-s^6 + \omega_c^6)} \tag{A-7}$$

The closed loop part of Figure V-1A may be redrawn as shown in Figure V-2A

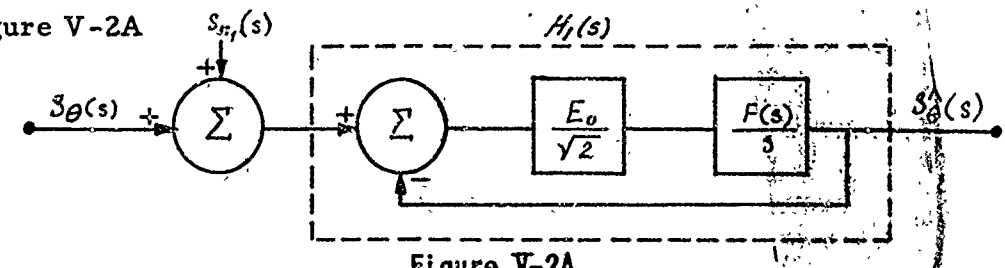


Figure V-2A

The filter which results in the minimum mean-square loop error is determined in terms of the power spectral densities of the input signal and noise by[11]

$$H_1(s) = \frac{\left[\frac{S_\theta(s)}{S_\pi^-(s)}\right]_+}{S_\pi^+(s)} \tag{A-8}$$

where $S_\pi(s) = S_\pi^+(s)\, S_\pi^-(s) = S_\theta(s) + S_n(s)$ (A-9)

and $S_\pi^+(s)$ possesses all the poles and zeroes of $S_\pi(s)$ that lie in the left half-plane and none of those that lie in the right half-plane.

The symbol $[\quad]_+$ implies that the function $\left[\frac{S_\theta(s)}{S_\pi^-(s)}\right]$ has been expanded into the form $\left[\frac{S_\theta(s)}{S_\pi^-(s)}\right]_+ + \left[\frac{S_\theta(s)}{S_\pi^-(s)}\right]_-$ where the first term has only left half-plane poles and the second has only right half-plane poles. It follows that the inverse Laplace transform of $\left[\frac{S_\theta(s)}{S_\pi^-(s)}\right]_+$ is equal to zero for $t < 0$, and $\mathcal{L}^{-1}\left\{\left[\frac{S_\theta(s)}{S_\pi^-(s)}\right]_-\right\}$ is zero for $t > 0$.

$$S_\theta(s) = \frac{A_0^2 \omega_c^6}{-s^2(-s^6+\omega_c^6)}$$

$$S_\pi(s) = \frac{2N_0}{E_0^2}\left[\frac{s^8 - \omega_c^6 s^2 + \frac{A_0^2 E_0^2}{2N_0}\omega_c^6}{-s^2(-s^6+\omega_c^6)}\right] \tag{A-10}$$

Let $$\alpha(s)\alpha(-s) = s^8 - \omega_c^6 s^2 + \frac{A_0^2 E_0^2}{2N_0}\omega_c^6 \tag{A-11}$$

and $$\beta(s)\beta(-s) = -s^2(-s^6+\omega_c^6) \tag{A-12}$$

$$S_\pi^+(s) = \sqrt{\frac{2N_0}{E_0^2}}\,\frac{\alpha(s)}{\beta(s)} \quad \& \quad S_\pi^-(s) = \sqrt{\frac{2N_0}{E_0^2}}\,\frac{\alpha(-s)}{\beta(-s)} \tag{A-13}$$

Upon substituting in Equation (A-8) one obtains

$$H_1(s) = \frac{\left[\dfrac{\dfrac{A_0^2 \omega_c^6}{\beta(s)\beta(-s)}}{\sqrt{\dfrac{2N_0}{E_0^2}}\dfrac{\alpha(-s)}{\beta(-s)}}\right]_+}{\sqrt{\dfrac{2N_0}{E_0^2}}\dfrac{\alpha(s)}{\beta(s)}} = \frac{\left[\dfrac{A_0^2 \omega_c^6}{\sqrt{\dfrac{2N_0}{E_0^2}}\beta(s)\alpha(-s)}\right]_+}{\sqrt{\dfrac{2N_0}{E_0^2}}\dfrac{\alpha(s)}{\beta(s)}} \tag{A-14}$$

$$\alpha(s)\alpha(-s) = s^8 - \omega_c^6 s^2 + \frac{A_0^2 E_0^2}{2N_0}\omega^6 = -s^2(-s^6 + \omega_c^6) + \frac{A_0^2 E_0^2}{2N_0}\omega_c^6$$

$$= \beta(s)\beta(-s) + \frac{A_0^2 E_0^2}{2N_0}\omega_c^6 \tag{A-15}$$

$$\frac{A_0^2 E_0^2}{2N_0}\omega_c^6 = \alpha(s)\alpha(-s) - \beta(s)\beta(-s)$$

then

$$H_1(s) = \frac{\left[\dfrac{\alpha(s)\alpha(-s) - \beta(s)\beta(-s)}{\beta(s)\alpha(-s)}\right]_+}{\dfrac{\alpha(s)}{\beta(s)}} = \frac{\left\{\left[\dfrac{\alpha(s)}{\beta(s)} - 1\right] - \left[\dfrac{\beta(-s)}{\alpha(-s)} - 1\right]\right\}_+}{\dfrac{\alpha(s)}{\beta(s)}} \tag{A-16}$$

$$= 1 - \frac{\beta(s)}{\alpha(s)} \tag{A-17}$$

$\alpha(s)$ may be determined by use of Ferrari's solution of the general quartic equation, and $\beta(s)$ by use of Tartaglia's solution of the general cubic equation[12]

We have, from Equation (A-11)

$$\alpha(s)\alpha(-s) = s^8 - \omega_c^6 s^2 + \delta^2 \omega_c^6$$

where

$$\delta = \sqrt{\frac{A_0^2}{\frac{2N_0}{E_0^2}}}$$

$\alpha(s)$ contains all the realizable roots, i.e., all roots in the left half-plane and $\alpha(-s)$ contains all the roots in the right half-plane. The roots in the right half-plane are mirror images of the roots in the left half-plane, thus, by finding the roots of s^2 and then taking the square root, the desired factorization of the roots can be obtained. Applying Ferrari's method, we obtain

$$\alpha(s) = s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0$$

$$a_0 = \delta\omega_c^3$$

$$a_1 = \chi\sqrt{2\chi}\left\{\sqrt{2+\gamma}\sqrt{\sqrt{2-\gamma}+1} + \sqrt{2-\gamma}\sqrt{\sqrt{2+\gamma}-1}\right\}$$

$$a_2 = \chi\left\{\sqrt{2-\gamma}+\sqrt{2+\gamma} + 2\sqrt{\sqrt{4-\gamma^2}-1} + \sqrt{2+\gamma}-\sqrt{2-\gamma}\right\}$$

$$a_3 = \sqrt{2\chi}\left\{\sqrt{\sqrt{2-\gamma}+1} + \sqrt{\sqrt{2+\gamma}-1}\right\}$$

$$\chi = \sqrt{\frac{\delta\omega_c^3}{\sqrt{3}}}\cos\theta/3$$

$$\gamma = \sqrt{4 - 3\sec^2\theta/3} = \frac{1}{\cos\theta/3}\sqrt{4\cos^2\theta/3 - 3}$$

$$\theta = \cos^{-1}\frac{3\sqrt{3}\omega_c^3}{16\delta^3}$$

From Equation (A-12) and using Tartaglia's method, we obtain

$$\beta(s) = s^4 + 2\omega_c s^3 + 2\omega_c^2 s^2 + \omega_c^3 s$$

$$= s^4 + b_3 s^3 + b_2 s^2 + b_1 s$$

Thus the loop filter is

$$F(s) = \frac{\sqrt{2}}{E_o} \quad \frac{s(\alpha(s)-\beta(s))}{\beta(s)}$$

The next step is to determine the output filter G(s). The model of the linearized phase-locked loop can be redrawn in the following form:

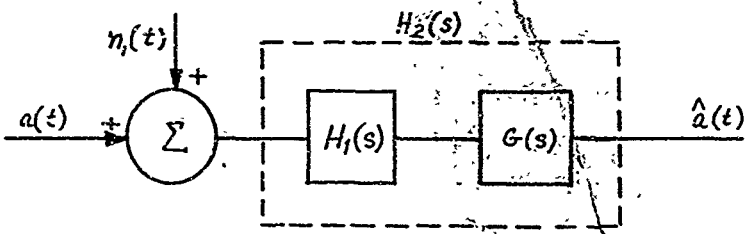


Figure V-3A

where

$$S_a(s) = \frac{A_o^2 \omega_c^6}{-s^6 + \omega_c^6}$$

and

$$S_n(s) = -s^2 \left[\frac{2N_o}{E_o^2}\right]$$

$$H_2(s) = H_1(s)\, G_1(s)$$

$$S_r(s) = S_a(s) + S_n(s) = -s^2\left[\frac{2N_o}{E_o^2}\right] \frac{\alpha(s)\alpha(-s)}{\beta(s)\beta(-s)}$$

$$H_2(s) = \frac{\left[\frac{S_a(s)}{S_r^-(s)}\right]_+}{S_r^+(s)}$$

$$H_2(s) = \frac{\left\{ \frac{\frac{A_o^2}{\frac{2N_o}{E_o^2}}\omega_c^6}{\frac{\alpha(s)\alpha(-s)}{\sqrt{\frac{2N_o}{E_o^2}}\, s\, \frac{\alpha(s)}{\beta(s)}}} \right\}_+}{\sqrt{\frac{2N_o}{E_o^2}}\, s\, \frac{\alpha(s)}{\beta(s)}}$$

$$\frac{\frac{A_o^2}{2N_o}}{\frac{E_o^2}{E_o^2}} \omega_c^6 = \alpha(s)\alpha(-s) - \beta(s)\beta(-s)$$

$$H_2(s) = \frac{\left\{\sqrt{\frac{2N_o}{E_o^2}}\left[1 - \frac{\beta(s)\beta(-s)}{\alpha(s)\alpha(-s)}\right] s\,\frac{\alpha(s)}{\beta(s)}\right\}_+}{\sqrt{\frac{2N_o}{E_o^2}}\; s\;\frac{\alpha(s)}{\beta(s)}}$$

$$= \frac{\left\{s\,\frac{\alpha(s)}{\beta(s)} - s\,\frac{\beta(-s)}{\alpha(-s)}\right\}_+}{s\,\frac{\alpha(s)}{\beta(s)}}$$

$$= \frac{\left\{s\left[\frac{\alpha(s)}{\beta(s)} - 1\right] - s\left[\frac{\beta(-s)}{\alpha(-s)} - 1\right]\right\}_+}{s\,\frac{\alpha(s)}{\beta(s)}}$$

$$\alpha(s) = s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_o$$

$$\beta(s) = s^4 + b_3 s^3 + b_2 s^2 + b_1 s$$

The term $s\left[\frac{\alpha(s)}{\alpha(-s)} - 1\right]$ contains only realizable parts.

The term $s\left[\frac{\beta(-s)}{\alpha(-s)} - 1\right]$ contains realizable and nonrealizable parts, but by expanding and dividing, the realizable part can be extracted.

$$-s\left[\frac{\alpha(-s)-\beta(-s)}{\alpha(-s)}\right]$$

$$=(-s)\left\{\frac{\left[a_3-b_3\right](-s)^3+\left[a_2-b_2\right](-s)^2+\left[a_1-b_1\right](-s)+a_0}{(-s)^4+a_3(-s)^3+a_2(-s)^2+a_1(-s)+a_0}\right\}$$

$$=\frac{\left[a_3-b_3\right](-s)^4+\left[a_2-b_2\right](-s)^3+\left[a_1-b_1\right](-s)^2+a_0(-s)}{(-s)^4+a_3(-s)^3+a_2(-s)^2+a_1(-s)+a_0}$$

$$=\left[a_3-b_3\right]+\text{nonrealizable parts}$$

Then $$H_2(s)=\frac{s\left[\frac{\alpha(s)}{\beta(s)}-1\right]-\left[a_3-b_3\right]}{s\frac{\alpha(s)}{\beta(s)}}$$

$$H_2(s)=1-\frac{\beta(s)}{\alpha(s)}\left[1+\frac{\left[a_3-b_3\right]}{s}\right]$$

$$H_1(s)=1-\frac{\beta(s)}{\alpha(s)}$$

From Equation (A-2)

$$\frac{E_o}{\sqrt{2}}F(s)=s\frac{H_1(s)}{1-H_1(s)}=s\frac{\left[\alpha(s)-\beta(s)\right]}{\beta(s)}$$

$$H_2(s)=H_1(s)\,G_1(s)$$

$$G_1(s)=\frac{H_2(s)}{H_1(s)}$$

$$G(s) = \frac{\dfrac{s\alpha(s) - \beta(s)\left[s + (a_3 - b_3)\right]}{s\,\alpha(s)}}{\dfrac{\alpha(s) - \beta(s)}{\alpha(s)}}$$

$$= \frac{s\left[\alpha(s) - \beta(s)\right] - (a_3 - b_3)\beta(s)}{s\left[\alpha(s) - \beta(s)\right]}$$

REFERENCES

[1] Lawton, John G. "Investigation of Analog and Digital Communication Systems" (Phase 3 Report) RADC-TDR-63-147, CAL Report No. UA-1420-S-3, May 1963. (AD No. 407 343)

[2] Akima, Hiroshi, "Theoretical Studies on Signal-to-Noise Characteristics of an FM System", IEEE Transactions on Space Electronics and Telemetry, Vol. SET-9, No. 4, December 1963, pp. 101-108.

[3] Youla, D.C. "The Use of the Method of Maximum Likelihood in Estimating Continuous-Modulated Intelligence Which Has Been Corrupted by Noise", IRE Transactions PGIT-3, March 1954, pp. 90-105.

[4] Becker, Harold D., Chang, Ting T., and Lawton, John G. "Investigations of Advanced Analog Communications Techniques", RADC-TDR-65-81, CAL Rpt. No. CM-1895-S-1, February 1965. (AD No. 613 703)

[5] Abramson, Norman, "Bandwidth and Spectra of Phase-and-Frequency-Modulated Waves", IEEE Transactions on Communication Systems, December 1963.

[6] Schilling, Donald L. and Billig, Joseph, "A Comparison of the Threshold Performance of the Frequency Demodulator using Feedback and the Phase-Locked Loop", Research Rpt. PIBMRI-1207-64, Polytechnic Institute of Brooklyn, February 28, 1964.

[7] Skinner, F.J. "Radio Transmission Systems - Theoretical Noise Performance Curves for Frequency Modulation Receivers Operating Below the Breaking Region", Bell Telephone Lab, unpublished memo, File 36690-1, 1 February 1954.

[8] Splitt, Frank G. "Design and Analysis of a Linear Phase-Locked Loop of Wide Dynamic Range", IEEE Transactions on Communication Technology, August 1966.

[9] Robinson, L.M. "Tanlock: A Phase-Lock Loop of Extended Tracking Capability", Proc. of National Winter Conference on Military Electronics, 1962.

[10] Frankle, J. "Threshold Performance of Analog FM Demodulators", RCA Review, December 1966, pp. 521-562.

[11] Newton, Gould, and Kaiser, Analytical Design of Linear Feedback Controls (see p. 154) John Wiley & Sons, Inc., New York 1961.

[12] Brink, R.W. Algebra College Course , D. Appleton-Century Company, Inc., New York, 1934.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Cornell Aeronautical Laboratory, Inc.
Buffalo, New York 14221

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Optimum Demodulation of PM and FM Signals

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Final Report (14 April 1966 - 27 May 1967)

5. AUTHOR(S) (First name, middle initial, last name)
Dr. John G. Lawton
Mr. Harold D. Becker
Dr. John T. Fleck
Mr. Sol Kaufman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1967	184	34

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF30(602)-4221	
b. PROJECT NO. 4519	UB-2254-B-1
c. Task 451902	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	RADC-TR-67-335

10. DISTRIBUTION STATEMENT
This document is subject to special export controls and each transmittal to foreign governments, foreign nationals or representatives thereto may be made only with prior approval of RADC (EMCI), GAFB, N.Y. 13440.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	RADC (EMCRS) Griffiss AFB NY 13440

13. ABSTRACT

This report presents results of investigating the limiting performance attainable with angle modulation (PM and FM) receivers. Statistical estimators which implement the criteria of minimum mean-square-error (MMSE) and maximum a posteriori probability (MAP) are compared with phase-locked loops. Mathematical investigations were carried out for the threshold region. Asymptotic formulae were developed where applicable; the region which could not be effectively treated analytically was analyzed by using a digital computer. The threshold phenomenon of phase-locked loops was investigated by means of a hybrid computer simulation. The results indicate that all properly designed demodulators for angle-modulation signals attain the same improvement above threshold, and that the major difference in demodulation techniques lies in the location of the threshold. The limitations of the FM threshold are explicitly brought out, and the factors which limit the extension of present-day threshold are pointed out. The maximum amount of threshold extension which it is theoretically possible to attain is determined.

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Frequency Modulation Statistical Demodulation Communication Information Theory Mathematical Optimization						